UNIVERSITY OF LIVERPOOL

The University of Liverpool

VALIDATION OF NOVEL COLORECTAL CANCER BIOMARKERS DERIVED FROM ANIMAL MODELS OF *Apc* INACTIVATION: ANALYSIS OF COHORTS FROM THE UK AND BRAZIL.

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Cleberson Jean dos Santos Queiroz

March 2017

ACKNOWLEDGEMENTS

Coming from a humble family from Brazil´s countryside, attending medical school was already an unimaginable achievement. Today, many years later, the opportunity of obtaining a PhD degree in the UK testifies how far we can go with hard work, dedication and, above all, support from others. This thesis was only possible due to the encouragement and support given by several outstanding persons, who I now have the honour to acknowledge.

First and foremost, I would like to thank my primary supervisor Dr John Jenkins who took the risks of accepting an unknown foreign student with very limited research background. John has given me a fantastic support in all stages of my research, guiding me from the initial laboratory training to the discussion of the final results. If I am now close to become an independent researcher, it is mostly because of him.

I would also like to thank my secondary supervisor Mark Pritchard. Professor Pritchard has helped to clarify the research pathway in confusing moments and has permitted me to work in special conditions without which I could not have completed this research. He also provided a clinical point of view that improved the design of the study and enriched the discussion of results.

In this Acknowledgements section, a special place is dedicated to my secondary supervisor Dr Fabio Miyajima. Fabio´s participation in my PhD preceded the start of this research by two years when he invited me and convinced me to accept the challenge of moving to a different country in order to embrace this doctoral project. He helped me during all these years not only providing scientific guidance but also helping me with personal issues. In fact, Fabio´s friendship is certainly one of the most important achievements of my PhD.

Another important person for this research was Veridiana Pessoa. Veri gave me invaluable help during laboratory experiments. Her expertise in cell culture and genomic techniques helped me to learn and succeed in

experiments with which I had no previous experience. Besides, Veri also took care of my family as only a careful mother would do and, for this, we are thankful forever.

I would like express my gratitude to the staff of the Gastroenterology Research Unit. Professor Barry Campbell helped me as a mentor giving tips and guidance. Mr Dave Barry was always available to help when I ran out of reagents. Dr Carrie Duckworth was always willing to lend her extensive knowledge when scientific questions seemed unsolvable.

Special thanks are directed to Dr Dale Vimalachandran for providing clinical samples from the Countess of Chester Hospital NHS Foundation Trust, and to Nadeem Al-Khafaji who performed some of the preliminary experiments that supported the research hypothesis leading to my work.

From Brazil, some colleagues also provided essential support. I would like to thank Dr Lenuce Ydy who helped me with the collection of the Brazilian samples and Dr Ivana Menezes who helped me providing samples and analysing pathology slides.

I would like to acknowledge the support given by the Brazilian government via the Science without Borders programme, an initiative that sponsored thousands of Brazilian research students abroad.

Lastly, I dedicate this thesis to my family. My beloved wife Jandi who bravely supported me, leaving our settled life, relatives and friends in Brazil to move to a different country with the responsibility of taking care of everything whilst I dedicated myself to this research. She has accomplished this task brilliantly. Moreover, during this time, she gave me the most powerful incentive possible: our daughter Alice, whose story is now eternally connected to my PhD and my time in the UK.

VALIDATION OF NOVEL COLORECTAL CANCER BIOMARKERS DERIVED FROM ANIMAL MODELS OF *Apc* INACTIVATION: ANALYSIS OF COHORTS FROM THE UK AND BRAZIL.

ABSTRACT

Colorectal cancer (CRC) is the third and second most common cause of cancer death in men and women, respectively, worldwide. Most deaths result from late diagnosis and the lack of effective treatments for patients with advanced disease. Better biomarkers for early diagnosis, prediction of response to treatment and prognostic determination are therefore urgently needed. In this research, we have assessed several CRC candidate biomarkers that had previously been identified during studies involving animal models of adenomatous polyposis coli *(Apc)* inactivation (the most common genetic alteration in colorectal carcinogenesis). Our hypothesis was that these candidate proteins would translate into valid biomarkers of human colorectal neoplasia. Therefore, we tested the expression of these candidate biomarkers in tissue and blood samples obtained from patients with colorectal neoplasia as well as healthy controls. Patient cohorts from the UK and Brazil were analysed in this research.

Using electronic scoring tools, we assessed the immunohistochemical expression of the candidate proteins in normal colonic mucosa, adjacent non-neoplastic colonic mucosa, colonic adenomas and colorectal cancer samples. Clear differential patterns of expression were observed for nucleosome assembly protein 1 – like 1 (NAP1L1), ribosomal protein L6 (RPL6) and prohibitin (PHB) when comparing cancers and non-malignant tissues. Additionally, NAP1L1 and RPL6 exhibited different expression patterns in low-grade versus high-grade adenomas, thus suggesting that they may play roles in the transition from low-risk to high-risk premalignant lesions.

Gene expression studies showed that *NAP1L1* and *RPL6* were highly expressed in the tumour and the adjacent mucosa from patients with CRC when compared to colonic biopsies obtained from normal control subjects. These results support a role for these genes not only in colorectal carcinogenesis but also in colonic “field cancerisation”. *RPL6* silencing resulted in strong inhibition of proliferation in HCT116 colorectal cancer cells. PCR-array studies demonstrated that *RPL6* silencing caused up-regulation of BCL2 associated X *(BAX)* and mutS homolog 2 (*MSH2*) - protectors against cancer development, and down-regulation of matrix metalloproteases 12 and 13 (*MMP-12* and *MMP-13*) - promoters of cancer progression, supporting the importance of *RPL6* in colorectal carcinogenesis.

The blood concentrations of NAP1L1 (assessed using a novel in-house electrochemiluminescence immunoassay), RPL6 and PHB (measured using commercial enzyme-linked immunosorbent assay kits) did not show any significant differences in cancer individuals when compared with normal controls and adenoma-bearing individuals. However, several new findings related to the measurement of the concentrations of these proteins in blood-derived fluids were made.

A study of a retrospective cohort of CRC patients clearly demonstrated that the immunohistochemical expression of NAP1L1 was related to prognosis. High nuclear expression of NAP1L1 was independently associated with a marked increase in overall survival and 5-year survival estimates. Mortality in this group was 61 to 72% lower when compared with the low-expression group. This difference was however only observed in patients who had late stage disease (stages III and IV).

The original contribution of this thesis is the confirmation that the candidate biomarkers derived from animal models of *Apc* inactivation are also differentially expressed in human CRC samples. The results produced by the various methodologies described suggest that NAP1L1, RPL6 and PHB may be potential novel biomarkers for the early diagnosis of CRC and the identification of high-risk premalignant lesions. Additionally, the association of NAP1L1 expression with the prognosis of CRC patients has not been previously reported and may have a clinical application. Further prospective research assessing larger sample cohorts is now highly recommended in order to confirm these findings.

VALIDATION OF NOVEL COLORECTAL CANCER BIOMARKERS DERIVED FROM ANIMAL MODELS OF *Apc* INACTIVATION: ANALYSIS OF COHORTS FROM THE UK AND BRAZIL.

TABLE OF CONTENTS:

List of abbreviations

95%CI –	95% confidence interval
ABC –	ATP-binding cassette
ABL –	Tyrosine-protein kinase transforming protein Abl
AKT –	AKT serine/threonine kinase 1
ANOVA –	Analysis of variance
APC –	Adenomatous polyposis coli
APES –	Aminopropyltriethoxysilane
ATP –	Adenosine triphosphate
AUC –	Area under the curve
B2M –	Beta-2-microglobulin
BAD –	Bcl2-associated agonist of cell death
BAK –	BCL2 antagonist/killer
BAT26 –	A repeat of 26 deoxyadenosines localized in an intron of *hMSH2*
BAX –	BCL2 associated X
BCL2 –	B-cell lymphoma 2
$BCLX_L$ –	B-cell lymphoma extra large
BCRP –	Breast cancer resistance protein
BID –	BH3 interacting domain death agonist
BIK –	BCL2 interacting killer
BMP3 –	Bone morphogenetic protein 3
BRAF –	B-Raf proto-oncogene
BRCA –	BRCa homolog tumor suppressor gene
CBL –	E3 ubiquitin-protein ligase CBL
CCSA-2 –	Cytochrome c biogenesis protein 2
CDC5L –	Cell division cycle 5-like
CDK –	Cyclin-dependent kinase
cDNA –	Complementary DNA
CDX –	Transcription factor protein
CEA –	Carcinoembryonic antigen
CIMP –	CpG island methylator phenotype
CIN –	Chromosomal instability
CK –	Cytokeratin
CLU –	Clusterin
CONEP –	"*Comissão Nacional de Ética em Pesquisa*" (National Commission for Research Ethics)
CpG –	Cytosine and guanine separated by one phosphate
CQ –	Cleberson Queiroz
CRC –	Colorectal cancer
CRK –	Adapter molecule CRK

CRMP-2 – Cysteine repeat modular protein 2
CT – Threshold cycle
CTNNB1 – β-catenin gene
CV – Coefficient of variation
DAB – 3,3’-diaminobenzidine
DAF – Decay-accelerating factor
DDX5 – DEAD box protein 5
DF-5 – Cytochrome oxidase subunit I (from *Demodex folliculorum*)
DISC – Death-inducing signalling complex
DLL4 – Delta-like ligand 4
DMSO – Dimethyl sulfoxide
DNA – Deoxyribonucleic acid
DNMT1 – DNA methyltransferase 1
DVL – Dishevelled
E2F – Transcription factor E2F
ECL – Electrochemiluminescence
ECM – Extra cellular matrix
EDTA – Ethylenediamine tetraacetic acid
EGF – Epidermal growth factor
EGFR – Epidermal growth factor receptor
EGR1 – Early growth response protein 1
EIA – Enzyme immunoassay
ELISA – Enzyme-linked immunosorbent assay
ELK – ETS domain-containing protein Elk-1
ERB-B2 – Erb-b2 receptor tyrosine kinase 2
ERK – Extracellular regulated MAP kinase
FABP6 – Fatty acid binding protein 6
FAP – Familial adenomatous polyposis
Fas – Fas cell surface death receptor
FasL – Fas ligand
FDZ – Frizzled
FFPE – Formalin-fixed, paraffin-embedded
FGF – Fibroblast growth factor
FIT – Faecal immunological test
FOBT – Faecal occult blood test
FOS – Proto-oncogene c-Fos
FRET – Fluorescent resonance energy transfer
FS – Flexible sigmoidoscopy
GAP – GTPase activating protein
GAPDH – Glyceraldehyde-3-phosphate dehydrogenase
GDP – Guanosine-diphosphatase
GEF – Guanine nucleotide exchange factor
GRB2 – Growth factor receptor-bound protein 2

GRB7 –	Growth factor receptor-bound protein 7
GSK-3 –	Glycogen synthase kinase *3*
GTP –	Guanosine-5'-triphosphate
HER-2 –	Human epidermal growth factor receptor 2
HMGB1–	High mobility group box 1
HNPCC –	Hereditary non-polyposis colorectal cancer
HPRT –	Hypoxanthine guanine phosphoribosyl transferase
HR –	Hazard ratio
HRAS –	Harvey rat sarcoma proto-oncogene
IGFBP2 –	Insulin like growth factor binding protein 2
IHC –	Immunohistochemistry
Int-1 –	Integrase 1
IPA –	Ingenuity pathway analysis
IPO8 –	Importin 8
JNK –	c-Jun NH(2)-terminal kinase
JNKK –	MAP2K4 mitogen-activated protein kinase kinase 4
JUN –	Transcription factor AP-1
Ki67 –	Cell proliferation antigen Ki-67
KRAS –	Kirsten rat sarcoma proto-oncogene
LC-MS/MS –	Liquid chromatography tandem mass spectrometry
LEF –	Lymphoid enhancer factor
LOH –	Loss of heterozygozity
M2-PK –	M2 pyruvate kinase
MALDI-TOF –	Matrix-assisted laser desorption ionisation time-of-flight
MAPK	Mitogen-activated protein kinase
MCC –	Mutated in colorectal cancers
MCM –	MCM ATPase
MDM2 –	Murine double minute 2
MDR –	Multidrug resistance
MEK –	MAP kinase-ERK kinase
MGMT –	O-6-methylguanine-DNA methyltransferase
MIQE –	Minimum information for publication of quantitative real-time PCR experiments
miR –	Micro RNA
miRNA –	Micro RNA
MLH –	MutL homolog
MMP –	Matrix metalloprotease
MMR –	Mismatch repair
mRNA –	Messenger RNA
MRP –	MDR-associated protein
MSD –	Meso scale discovery
MSH –	MutS homolog
MSI –	Microsatellite instability
MSS –	Microsatellite stability

MTHFR – Methylenetetrahydrofolate reductase
mTOR – Mammalian target of rapamycin
MUC1 – Mucin 1
MYC – Myelocytomatosis oncogene
NA – Nadeem Al-Khafaji
NAP1L1 – Nucleosome assembly protein 1 – like 1
NCBI – National centre for biotechnology information
NCK – NCK adaptor protein 1
NCL – Nucleolin
NDRG4 – N-Myc downstream regulated family member 4
NET – Neuroendocrine tumour
NOXA – Phorbol-12-myristate-13-acetate-induced protein 1
NRAS – Neuroblastoma rat sarcoma proto-oncogene
NURD – Nucleosome remodeling deacetylase
ONS – Office for national statistics
PAK – p21-activated kinase
PBS – Phosphate buffered saline
PCR – Polymerase chain reaction
PDGF – Platelet derived growth factor
PHB – Prohibitin
PI3K – Phosphatidylinositol-4,5-bisphosphate 3-kinase
PIK3CA – Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha gene
PKC – Protein kinase C
PLCγ – Phosphatidylinositol-specific phospholipase C subunit γ
PMS – Post meiotic segregation
PPIA – Peptidylprolyl-isomerise A
PPV – Positive predictive value
PTEN – Phosphatase and tensin homolog
PTGS2 – Prostaglandin-endoperoxide synthase 2
PUMA – p53-upregulated modulator of apoptosis
qPCR – Quantitative PCR
RAC – Ras-related C3 botulinum toxin substrate
RAF – RAF proto-oncogene serine/threonine-protein kinase
RAGE – Receptor for advanced glycation end products
RAS – Rat sarcoma family of proto-oncogenes
RB – Retinoblastoma
RECQL4 – RecQ like helicase 4
RIA – Radioimmunoassay
RNA – Ribonucleic acid
ROC – Receiver operating characteristics
RP – Ribosomal protein
RPL6 – Ribosomal protein L6
RRM1 – Ribonucleotide reductase catalytic subunit M1

rRNA –	Ribosomal RNA
RT –	Reverse transcription
RTK –	Receptor tyrosine kinase
RUNX –	Runt-related transcription factor 1
S6K –	Ribosomal protein S6 kinase
SC35 –	Serine/arginine-rich splicing factor-like protein, putative
sCD26 –	Soluble form of CD26
SD –	Standard deviation
SDHA –	Succinate dehydrogenase complex flavoprotein subunit A
SDS-PAGE –	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE –	Standard error
SELDI-TOF –	Surface-enhanced laser desorption/ionization time-of-flight
SFRS2 –	Splicing factor arginine/serine-rich 2
SHC –	SHC-transforming protein 1
SHP2 –	K-box region and MADS-box transcription factor family protein
siRNA –	Small interfering RNA
SMAD –	Homologs of both the *Caenorhabditis elegans* protein SMA and the *Drosophila* protein MAD
SOS –	Son of sevenless
SP1 –	Sp1 transcription factor
SRB –	Sulforhodamine B
SRC –	SRC proto-oncogene, non-receptor tyrosine kinase
SSRP1 –	Structure specific recognition protein 1
SUMO –	Small ubiquitin-like modifier
SUP16H –	Suppressor of Ty 16 homolog
SUS –	"*Sistema Unico de Saude*" (Unified Health System)
TBS –	Tris-buffered saline
TCF –	T-cell factor
TGF-β –	Transforming growth factor β
TIMP-1 –	TIMP metallopeptidase inhibitor 1
TLR –	Toll-like receptor
TMA –	Tissue microarray
TNF –	Tumour necrosis factor
TNFR –	Tumour necrosis factor receptor
TNM –	Tumour-Node-Metastasis
TP53 –	Tumour protein 53
TRAIL –	TNF-related apoptosis-inducing ligand
TRIM28 –	Tripartite motif containing 28
UBC –	Polyubiquitin-C
UK –	United Kingdom

UNG –	Uracil-N glycosylase
USA –	United States of America
VAV –	Vav guanine nucleotide exchange factor
VEGF –	Vascular endothelial growth factor
WB –	Western blot
Wnt –	Gene named after the fusion of the gene names wingless and integrase1

List of tables

List of figures

Chapter One:

Introduction

1. CHAPTER 1 – INTRODUCTION

1.1. Epidemiology

1.1.1. Cancer - general aspects

Cancer is a major health problem worldwide. In 2012, 14.1 million new cases were estimated causing 8.2 million deaths (Ferlay *et al*., 2015). These figures established cancer as the second leading cause of death, but it becomes the leading cause when only high-income countries are taken into account (Mathers *et al*., 2008). Although it highlights the importance of cancer for developed countries, it should be noted that 64% of those deaths occurred in developing nations (Jemal *et al*., 2011). It has been predicted that cancer deaths will increase steadily in the next decades, as shown in figure 1.1 (World Health Organization, 2008).

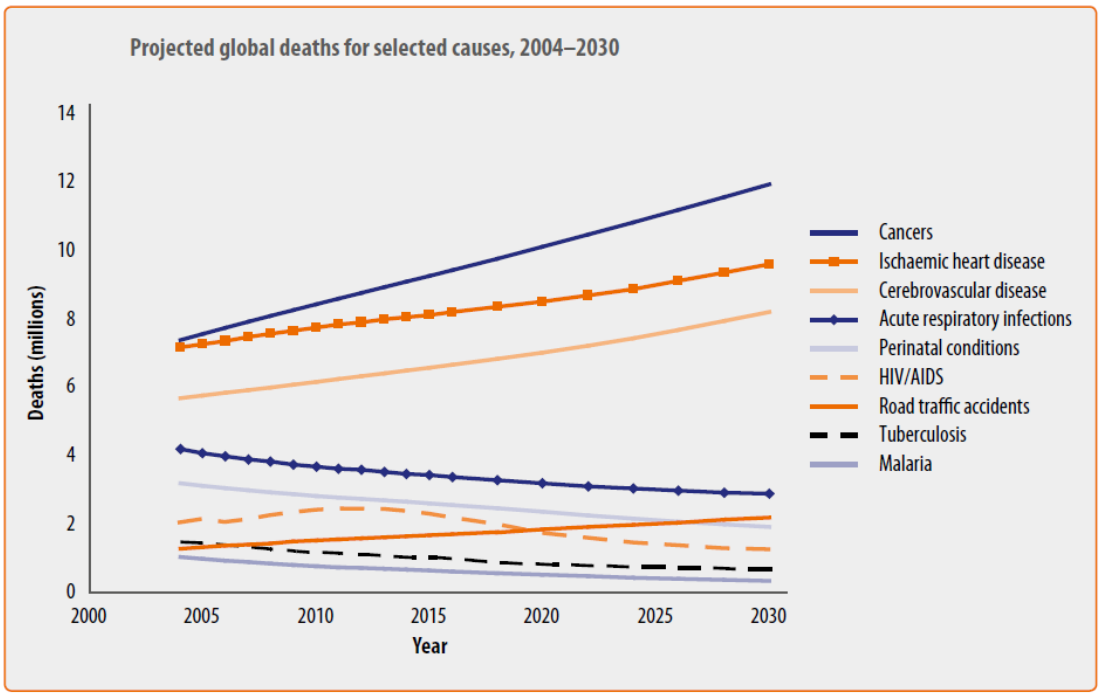


Figure 1.1. Projected global deaths for selected causes, 2004-2030. From (World Health Organization, 2008).

Cancer incidence and mortality vary widely in different regions. In a report of data from all continents, the overall cancer incidence ranged from 126/100,000 in African men to 398.4/100,000 in North American men (world average 209.6/100,000 men) (Kamangar *et al*., 2006). In most developed countries, although cancer incidence is increasing, overall mortality is decreasing. In the European Union, evaluating the period between 1990 and 2004, the cancer mortality rate diminished from 185.2 to 168/100,000 in men and from 104.8 to 96.4/100,000 in women (La Vecchia *et al*., 2010). The same trend was seen in North America (Kamangar *et al*., 2006) and Japan (Katanoda *et al*., 2013). Another noteworthy indicator demonstrating the contrast between regions is the *mortality-to-incidence ratio* (mortality ratio divided by incidence ratio, MR:IR). MR:IR approaching 1.0 suggests a limited survival and can reflect disease aggressiveness, late diagnosis or inadequate access to proper treatment. MR:IR was shown to range from 0.38 and 0.43 for North American and Oceanic men to 0.83 for African men. Europe and Central/South America have remained at an intermediate position (0.63 for both) (Kamangar *et al*., 2006).

1.1.2. **Colorectal cancer in the world**

Colorectal cancer (CRC) is the third most common type of cancer in both sexes, and the third and second most common cause of cancer death in men and women, respectively (Ferlay *et al*., 2015). Worldwide, new cases totalled 1.36 million in 2012, and 694,000 deaths occurred in the same year translating into an incidence rate of 17.2/100,000 and a mortality rate of 8.4/100,000 (Ferlay *et al*., 2015). Incidence of CRC varied 10-fold when comparing North America/Western Europe/Australasia (30.1 to 44.8/100,000 for men and 22.7 to 32.2/100,000 for women) to Western Africa (4.5 and 3.8/100,000 for men and women, respectively). Again, the *mortality-to-incidence ratio* was higher in less developed regions. MR:IR ranged from 0.34 in North America to 0.89 in Africa (Kamangar *et al*., 2006).

Besides these static figures, it is important to assess the trends in CRC epidemiology. With the globalisation of habits and westernisation of many cultures, changes in CRC incidence are occurring in some countries (Center *et al*., 2009). While most developed countries are experiencing a plateau or even a decrease in CRC incidence, developing areas are experiencing an increase. From 51 international cancer registries evaluated in one study, 27 showed an increase in CRC incidence, while only one (USA) showed a decrease for both males and females (Center *et al*., 2009). Selected Eastern Europe countries (such as Czech Republic, Slovakia, Slovenia, Poland and Estonia) exhibited a striking increase exceeding USA rates for men (see figure 1.2a-d). Further studies have shown increases in CRC incidence in Hong Kong (Xie *et al*., 2012), Spain (Bernal *et al*., 2009) and Serbia (Mihajlović *et al*., 2013). Other areas of increasing incidence include most parts of Asia and selected South American countries (Center *et al*., 2009). On the other hand, recent reports have demonstrated a reduced incidence in Japan (Katanoda *et al*., 2013) and Italy (Crocetti *et al*., 2010). Historically exhibiting a very high incidence of CRC, the reduction seen in the USA has also been demonstrated by other reports (Edwards *et al*., 2010, Siegel *et al*., 2012).

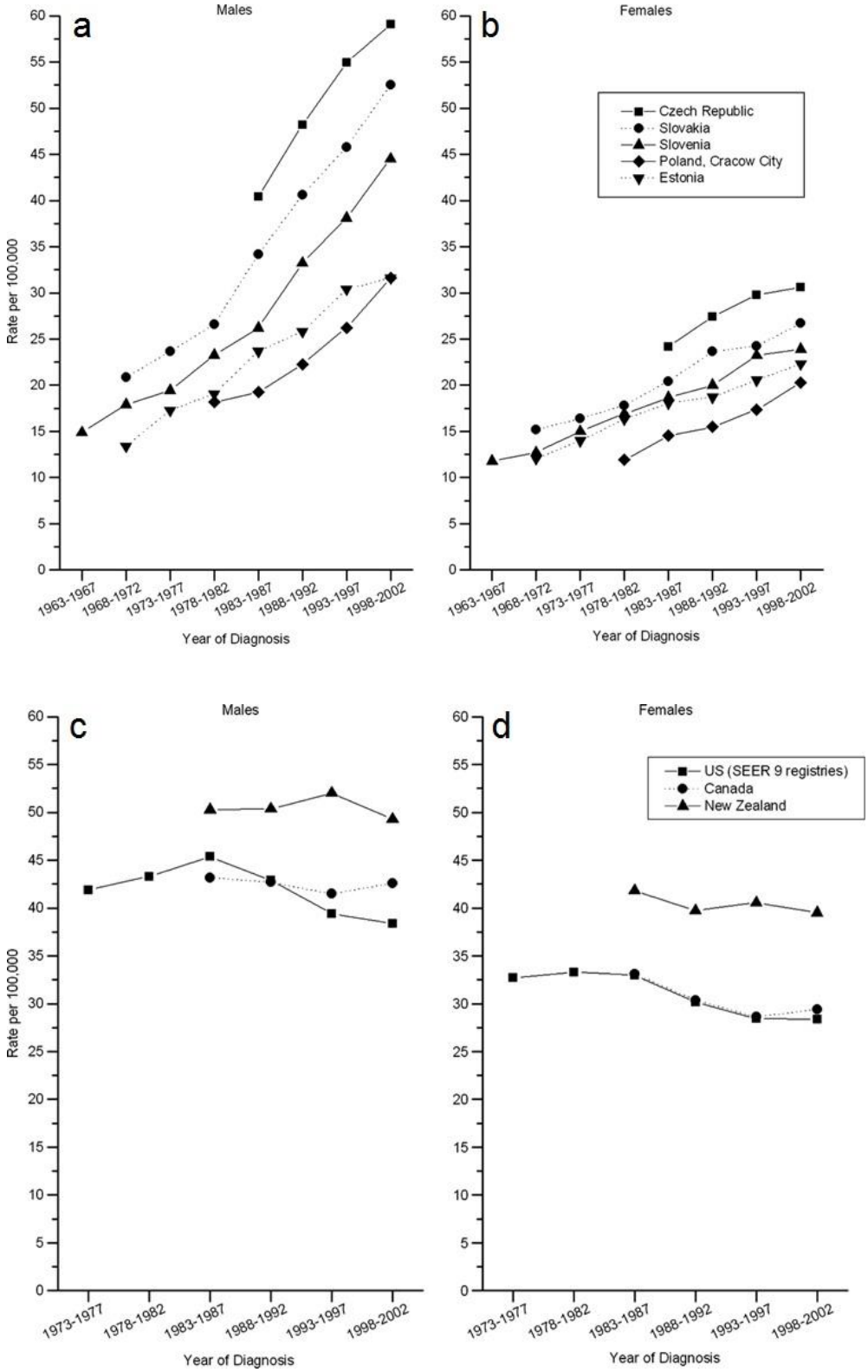


Figure 1.2. Colorectal cancer incidence in selected countries, according to gender. From (Center *et al*., 2009).

Focusing on CRC mortality, a similar pattern has been seen with the highest rates in Central/Eastern Europe (20.3 and 11.7/100,000 for men and women, respectively) and the lowest rates in Western Africa (3.5 and 3.0/100,000 for men and women, respectively) (Ferlay *et al*., 2015). Most developed countries have exhibited a decreasing trend in CRC mortality in recent decades (La Vecchia *et al*., 2010, Katanoda *et al*., 2013, Arfè *et al*., 2011, Edwards *et al*., 2010). Important exceptions in Europe include some Eastern countries such as Hungary and Czech Republic (La Vecchia *et al*., 2010) which demonstrated increasing mortality rates.

Striking differences in CRC epidemiology can be seen even when comparing European countries. Table 1.1 shows the incidence and mortality by CRC in selected countries. Spain has shown a transitional pattern in recent years. From 1975 to 2004, it exhibited a steady increase in CRC mortality (Bernal *et al*., 2009). After 2006, new figures and estimates indicated a trend to stabilisation or even a slight decrease both in incidence and mortality in males and females (Sanchez *et al*., 2010). In France, the incidence of CRC has increased by 0.99% in males and 0.83% in females annually since 1978 (Chevreul, 2010). In the same period, the mortality rate decreased by 0.76% in males and 1.07% in females every year. The overall 5 year-survival rate in France was 56%, ranging from 94% in stage I (early disease) to 5% in stage IV (advanced disease) (Chevreul, 2010). Recently, Rollot *et al*. analysed *net survival* (survival that might occur if cancer was the only cause of death) due to CRC in France since 1976. They reported an improvement in 5 year-*net survival* from 39% during 1976-1985 to 56% during 1996-2005 (Rollot *et al*., 2013). In the case of Germany, CRC incidence and mortality have decreased both for males and females (GEKID, 2013). The overall age-adjusted 5 year survival rate in Germany has improved from 60.6% in 2002 to 65% in 2006 (Majek *et al*., 2012). That improvement in survival was limited to local or regional disease stages, excluding advanced disease (stage IV). When sex differences were evaluated in patients with CRC in Germany, it was found that women had a higher 5 year survival rate compared to men (64.5% vs 61.9%) (Majek *et al*., 2013). That difference was most pronounced for patients less than 65 years

of age, which led the authors to hypothesise that sex hormones could play a role in that survival advantage via the regulation of immune or inflammatory responses.

Table 1.1. Incidence and mortality caused by CRC in selected European countries (per 100,000 age-adjusted population).

Incidence	**Males**	**Females**	**Reference**
Hungary	-	-	-
Czech Republic	59.1	30.6	(Center *et al*., 2009)
Spain	65.0	36.0	(Sanchez *et al*., 2010)
France	31.6 (combined)	-	(Chevreul, 2010)
Germany	58.7	36.9	(GEKID, 2013)
Italy (Parma Province)	42.6	27.1	(Center *et al*., 2009)
Mortality			
Hungary	34.6	18.2	(La Vecchia *et al*., 2010)
Czech Republic	35.8	17.9	(La Vecchia *et al*., 2010)
Spain	27.0	14.0	(Sanchez *et al*., 2010)
France	17.5	10.1	(La Vecchia *et al*., 2010)
Germany	22.3	13.9	(GEKID, 2013)
Italy	16.4	10.3	(Arfè *et al*., 2011)

1.1.3. **CRC in the United Kingdom (UK)**

According to the UK Office for National Statistics (ONS), 163,100 males and 159,800 females were diagnosed with cancer (all types) each year in the UK during 2008-2010 (Office for National Statistics, 2012). Of these, the average annual number of CRC cases was 22,517 in males and 17,864

in females, corresponding to an incidence of 58 new cases per 100,000 men and 37 new cases per 100,000 women. In the same period, an average of 8,569 men and 7,207 women died from CRC annually, which translated into mortality rates of 21/100,000 in men and 13/100,000 in women. A wide variation in CRC incidence has been demonstrated in the UK according to ethnic groups, with a higher occurrence amongst “whites” compared to “non-whites” (Ali *et al*., 2013). Recent data have shown a stabilisation in CRC incidence and a reduction in CRC mortality in England and in the UK (La Vecchia *et al*., 2010). As depicted in figure 1.3, an analysis of Southeast England’s population has demonstrated a steady decrease in CRC mortality since 1972 (Sanjoaquin *et al*., 2007).

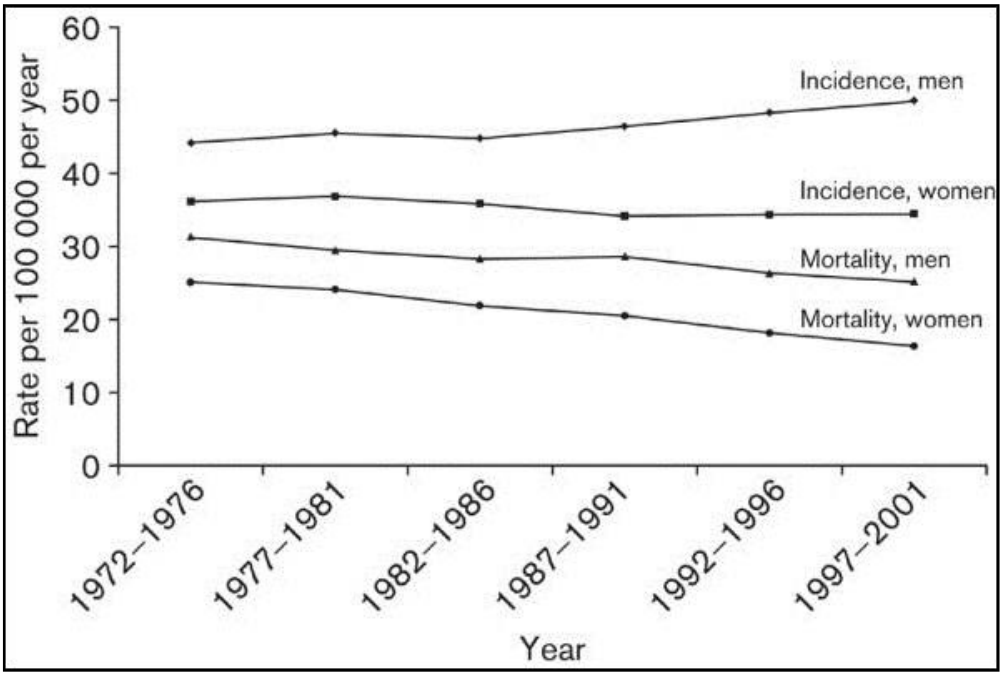


Figure 1.3. Colorectal cancer incidence and mortality in the Southeast England population. From (Sanjoaquin *et al*., 2007).

Although CRC mortality in the UK is declining, it is still higher than in some other European countries. When compared to Norway and Sweden, mortality due to CRC in England was higher, mainly due to deaths in the first 3 months after diagnosis (Morris *et al*., 2011). It was also demonstrated to be higher than in France (Dejardin *et al*., 2013), with most of the difference being attributed to early mortality especially within 1 year after diagnosis. A detailed analysis of this first year mortality due to CRC in England

demonstrated that most of the excess of deaths were confined to socioeconomically deprived groups (Moller *et al*., 2012).

1.2. The context of Brazil

Cancer is a disease with strong ethnic and geographic influence. As described above, there is an enormous variation in cancer incidence and mortality throughout the world (Crocetti *et al*., 2013, Li *et al*., 2013a, Kamangar *et al*., 2006). Besides, it has been demonstrated that some cancer biomarkers are present at different frequencies in different ethnic populations. For instance, epidermal growth factor receptor (EGFR)-activating mutations are recognised to be an important predictive marker in non-small cell lung cancer (Rosell *et al*., 2005). However, these particular mutations occur more frequently in Asian patients compared to their counterparts from Western countries (Shigematsu *et al*., 2005). Therefore, when studying cancer biomarkers, it is important to understand the ethnic background of the population under study in order to draw accurate conclusions about any result obtained, especially regarding the validity of those conclusions for other populations. The research described in this thesis involves analyses of samples obtained from a cohort of Brazilian patients with CRC.

Brazil is the 5th largest country in the world both by geographic area and population size. According to the 2010 national Census, it has almost 200 million inhabitants (Censo, 2010), which represents approximately 52% of the South American and 3% of the world's population (Giolo *et al*., 2012). Around five hundred years after the initial Portuguese colonisation, the composition of the Brazilian population is the result of important waves of migration between the 18th and 20th centuries, mainly by Europeans, Asians and Africans (slaves) (Levy, 1974). This diverse ethnic background makes the study of a Brazilian cohort attractive in terms of validating cancer biomarkers.

1.2.1. **Colorectal cancer in Brazil**

According to the Brazilian National Cancer Institute (INCA), the estimated incidence of CRC was 14.75 cases per 100,000 men and 15.95 cases per 100,000 women in 2012 (INCA, 2011). Those numbers are possibly underestimates since a large portion of the Brazilian population is not covered by population-based cancer registries and most of these data come from death certificates (Franca *et al*., 2008). Despite these drawbacks, several reports have clearly shown an increase in CRC incidence and mortality in Brazil in recent years. Assessing the mortality between 1980 and 2004, an ascending trend for both males and females has been demonstrated, as shown in figure 1.4 (Chatenoud *et al*., 2010). A similar trend was also described by Silva *et al*. (Silva *et al*., 2011). A recent publication reported that Brazil has had the greatest increase in CRC incidence amongst 184 countries analysed (Arnold *et al*., 2016). As a result of the continental size of the country, along with the remarkable variations in dietary habits and socio-economic conditions, mortality rates vary strikingly among Brazilian regions. An evaluation of the number of deaths due to CRC in different Brazilian state capitals has been performed (das Neves *et al*., 2005). The authors reported rates, per 100,000 persons, ranging from 1.2 (female) and 2.8 (male) in Teresina – Piauí state, to 10.4 and 11.8 in Porto Alegre – Rio Grande do Sul state. Since the former is one of the poorest Brazilian states and the latter is one of the wealthiest, it is supportive of a correlation between CRC rates and socio-economic status in the country. This notion is further supported by a study in which the authors found a clear relationship between increased income and higher CRC mortality in the Brazilian population (Guimarães *et al*., 2013).

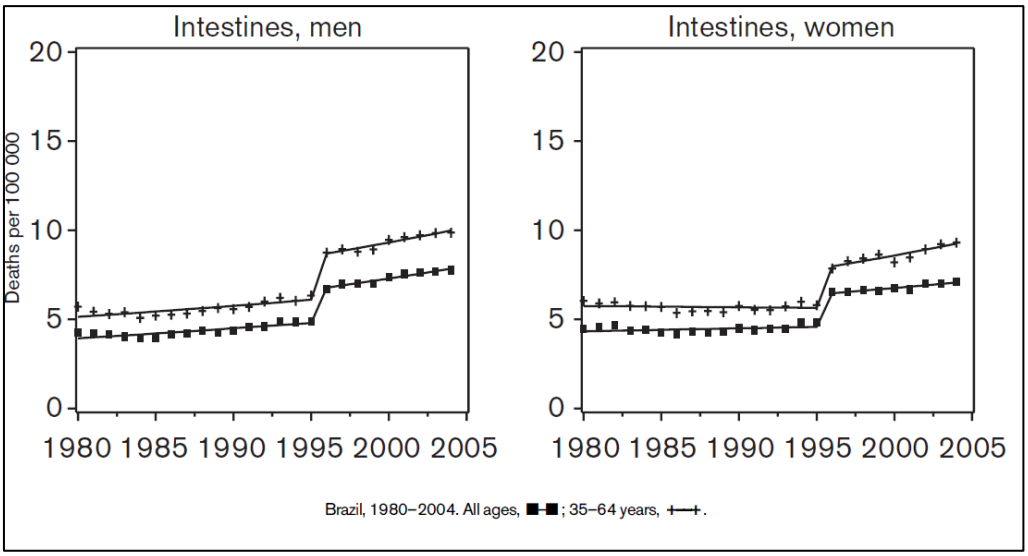


Figure 1.4. Trends in colorectal cancer mortality in Brazil, from 1980 to 2004. From (Chatenoud *et al*., 2010)

Another indicator of the increasing importance of CRC for the Brazilian Public Health System was given by Torres *et al*., who demonstrated that the costs associated with hospital admissions due to CRC increased more than 100% from 1996 to 2008 despite an overall decrease in the cost per admission (Torres *et al*., 2010). Although rising, CRC incidence and mortality in Brazil are still less than those seen in developed countries. This is explained, at least in part, by the delayed urbanisation and development of the country compared to the richest nations in the world. Nonetheless, in the last few decades the country has experienced a transformation due to economic growth and urbanisation, associated with a lifestyle closer to that seen in North America and Europe (Oliveira *et al*., 2010, Azevedo *et al*., 2008, Dumith *et al*., 2011). Besides this, longevity is also increasing in the country, so that a shift from a relatively young to an older population is occurring (Ferri, 2012). Hence, all data available to date indicate the possibility of a massive increase in CRC cases and its associated costs in Brazil over the next few decades unless effective strategies are put in place to prevent this occurrence.

1.3. The biology of colorectal cancer

1.3.1. Fundamentals of carcinogenesis

Cancer cells are characterised by uncontrolled growth and escape from anti-cancer defence mechanisms. In order to develop and progress, malignant tumours must modify the well-designed network that controls growth, division and the interaction with the environment (Weinberg, 2007). Since most of the effectors of those processes are proteins, it is intuitive to conclude that alterations in the activity of specific proteins are necessary to generate cells and tissues with those hallmarks. These proteins are produced by genes involved in the control of cell proliferation, survival and chromosomal stability. Such cornerstone genes can be grouped into two broad categories: oncogenes (or proto-oncogenes) and tumour-suppressor genes (Schmandt and Mills, 1993). Proto-oncogenes are genes which encode proteins involved in inducing cell proliferation. The majority of these proteins are transcription factors, chromatin modellers, growth factors, growth factor receptors, signal transducers, and apoptosis regulators (Croce, 2008). Consequently, proto-oncogenes play important roles in regulating the homeostasis of normal cells. When their function is abnormally activated (*gain-of-function* by chromosomal rearrangements, amplifications, mutations or epigenetic modifications), they are named oncogenes and, as a result of the permanent proliferative stimulus, confer a growth advantage or increased survival to the cells in which such alterations occur (Croce, 2008). An illustrative example is the *HER-2* gene (also termed *ERB-B2*), a member of the *epidermal growth factor receptor* family. The product of this gene, a tyrosine-kinase trans-membrane receptor, is normally expressed on the cell surface and is sensitive to the effects of growth factor ligands (Hung and Lau, 1999). In certain cases of cancer (such as some breast and gastric tumours), the gene encoding HER-2 undergoes amplification, which causes an increase in the expression of the receptor on the cell surface (Sellami *et al*., 1991, Marx *et al*., 2009). Therefore, cells are stimulated to proliferate even

under the influence of small amounts of growth factors, conferring more aggressiveness to these tumours (Kaptain *et al*., 2001).

On the other hand, tumour suppressors are genes responsible for halting cellular growth and proliferation in physiological settings. When inactivated (*loss-of-function*), tumour suppressors lose their controlling functions, allowing cells to replicate with no obstacle (Hansen *et al*., 1988). The first tumour suppressor gene cloned in humans was the Retinoblastoma gene (*RB*) (Friend *et al*., 1986). The product of this gene – the protein RB, controls the transition of cells from G1 to S-phase. Hence, its inactivation leads to uncontrolled transition to the phase of DNA synthesis. Mutations in *RB* were initially identified in patients suffering from a type of ocular cancer known as retinoblastoma (Classon and Harlow, 2002). Later, these mutations were found in several other types of cancer. Currently, different forms of inactivation of the RB pathway are thought to be present in virtually all cancers (Hahn and Weinberg, 2002, Sherr and McCormick, 2002), and many other functions have been attributed to the protein besides controlling G1/S transition (RB function will be further addressed below) (Rubin and Sage, 2013, Uchida, 2012). The list of oncogenes and tumour suppressors is extensive and beyond the scope of this thesis. These examples aim solely to provide a general overview of the concepts involved.

Although attractive, the notion that a unique mutation in either a proto-oncogene or a tumour suppressor is sufficient to cause a malignant phenotype is too simplistic. In fact, while alterations of these critical genes can be the initial step for cancer development, many more mutations are necessary for malignant transformation. This model, referred to as *multistep carcinogenesis*, predicts that, after an initial genetic defect that facilitates genomic instability and/or cellular replication, an accumulation of additional mutations must occur in order to enable a given cell to exhibit all the features of cancer (Yamada and Mori, 2007, Duesberg and Li, 2003). It has been demonstrated that about 80 mutations that alter amino acid sequences occur in a typical colorectal cancer (Wood *et al*., 2007). Statistical analysis has shown that only 15 or fewer of these 80 mutations are *drivers* involved in the initiation, progression or maintenance of the tumour. The majority of the

mutations are only *passengers* (harmless) and are accumulated during tumour progression (Wood *et al*., 2007). These *drivers* and *passengers* must be clearly differentiated when they are evaluated as candidate cancer biomarkers or drug targets.

Based on this concept of multistep carcinogenesis, Hanahan and Weinberg proposed a model to illustrate the constellation of capabilities which a given cell (and the tissue it forms) must acquire in order to develop the capacity of uncontrolled replication, progression, invasion and metastasis (Hanahan and Weinberg, 2011). These “hallmarks of cancer” are briefly described below:

a. Sustaining proliferative signalling

The cell cycle clock which ultimately controls cell proliferation is regulated by many factors. Eukaryotic cells have developed an extraordinarily complex network of signals that orchestrate cellular growth and division. Basically, this network is made up of extracellular factors (growth factors and inhibitors) and intracellular signalling pathways (Weinberg, 2007). The interaction between the extracellular and intracellular components is usually performed via trans-membrane proteins (receptors). Therefore, when a growth factor approaches the cell surface from the extracellular matrix, adjacent cells or the cell itself, it binds to the extracellular domain of the membrane receptor. The receptor then transmits a signal to the intracellular domain. Once in the cytoplasm, the signal continues travelling through a signalling cascade until it reaches its final destination (Hynes and MacDonald, 2009, Witsch *et al*., 2010, Lemmon and Schlessinger, 2010). There are many classes of growth factors/receptors and an even greater number of intracellular signalling pathways. The most important ones for colorectal carcinogenesis will be addressed later. As an example, figure 1.5 illustrates the *epidermal growth factor receptor* (EGFR) pathway. Therefore, in order to sustain proliferative signalling, cancer cells must develop autonomy in one or several pathways. This can be achieved through genetic alterations that lead to the sustained supply of growth factors

(Witsch *et al*., 2010), continuous activation of membrane receptors (Lemmon and Schlessinger, 2010, Audigier *et al*., 2013) or autonomy of critical points within the intracellular signalling network (Bell and Ryan, 2005).

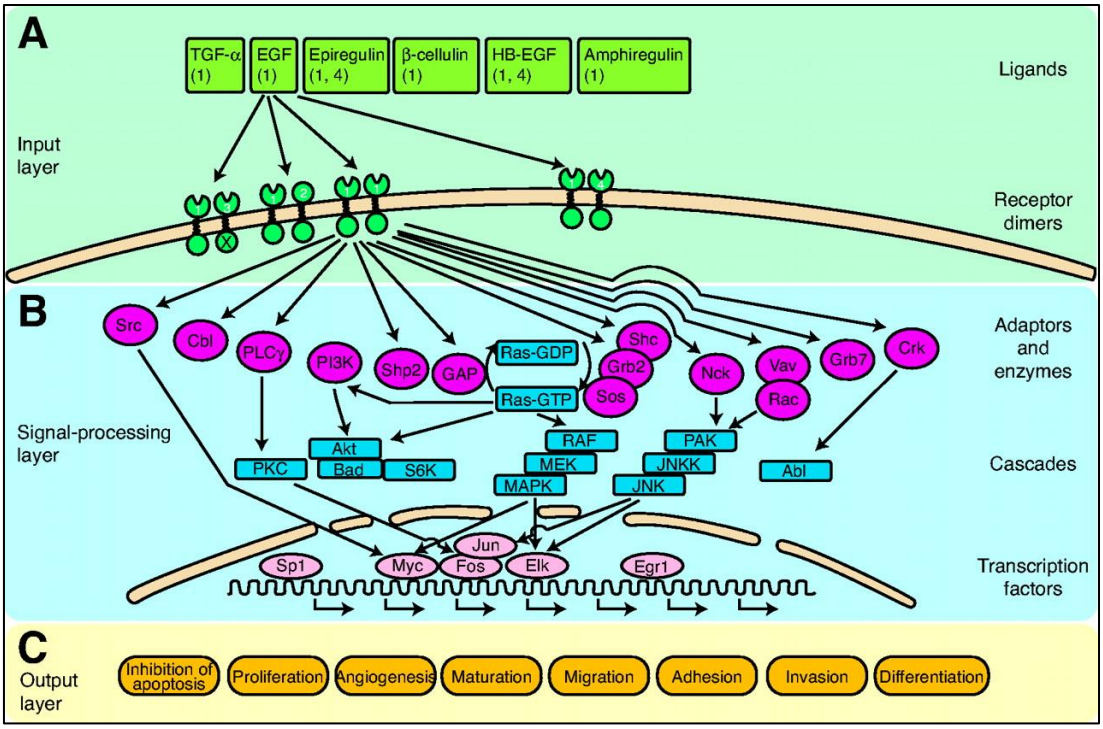


Figure 1.5. The EGFR signalling network. Ligands from the extracellular space bind to trans-membrane receptors. In this case, the signal is transmitted to the intracellular space through the activation of a tyrosine-kinase in the cytoplasmic domain of the receptor. This kinase phosphorylates downstream effectors triggering a signalling cascade which ultimately regulates the transcription of various genes linked to growth, proliferation and metastasis. Numbers in each ligand block indicate the respective high-affinity HER receptor among HER1 (1), HER2 (2), HER3 (3) and HER4 (4). HB-EGF, heparin-binding EGF. From (Arteaga, 2002), adapted from (Yarden and Sliwkowski, 2001).

b. Evading growth suppressors

Several stimulators of cell multiplication are counterbalanced by inhibitory mechanisms and the result of this equilibrium dictates whether the cell must proliferate or not (Weinberg, 2007). The protein RB plays a central role in the so called “cell cycle clock” (Dick and Rubin, 2013). It controls the cell’s transition from G1 to S-phase, which triggers DNA synthesis and prepares the cell for division (Classon and Harlow, 2002). Therefore, RB is

termed the gatekeeper of the cell cycle. However, RB determines the cell's fate based on various stimuli. As depicted in figure 1.6, RB works in conjunction with the E2F transcription factor family. When non-phosphorylated, RB binds to E2F preventing its translocation into the nucleus thereby inhibiting cell proliferation. Following a proliferative stimulus, RB is phosphorylated by cyclin proteins (cyclin D and E) and their cyclin-dependant kinase (CDKs) cognates, causing the release of E2F. Once free from RB attachment, E2F translocates into the nucleus where it activates the transcription of various pro-mitotic genes (Classon and Harlow, 2002).

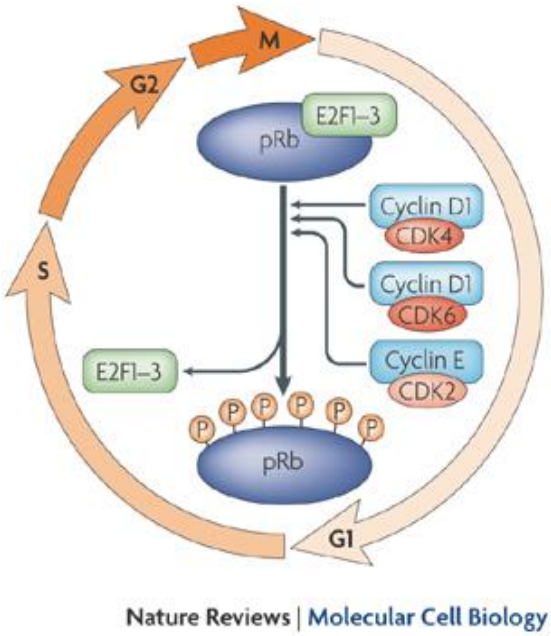


Figure 1.6. Interaction between Rb, E2F transcription factors and cyclins/CDKs during cell cycle progression. From (Coller, 2007).

Another important element of cell cycle control is p53. This protein is responsible for the surveillance of cells against conditions that cause threats to cell viability, especially DNA damage (Levine and Oren, 2009). In situations in which the threatening condition cannot be corrected, p53 activates several different pathways that can halt cellular growth (Baker *et al*., 1990a, Diller *et al*., 1990), induce apoptosis (Yonish-Rouach *et al*., 1991, Shaw *et al*., 1992) or senescence (Wang *et al*., 1998, Xue *et al*., 2007).

Figure 1.7 summarises some of the several mechanisms by which p53 responds to physiological stress.

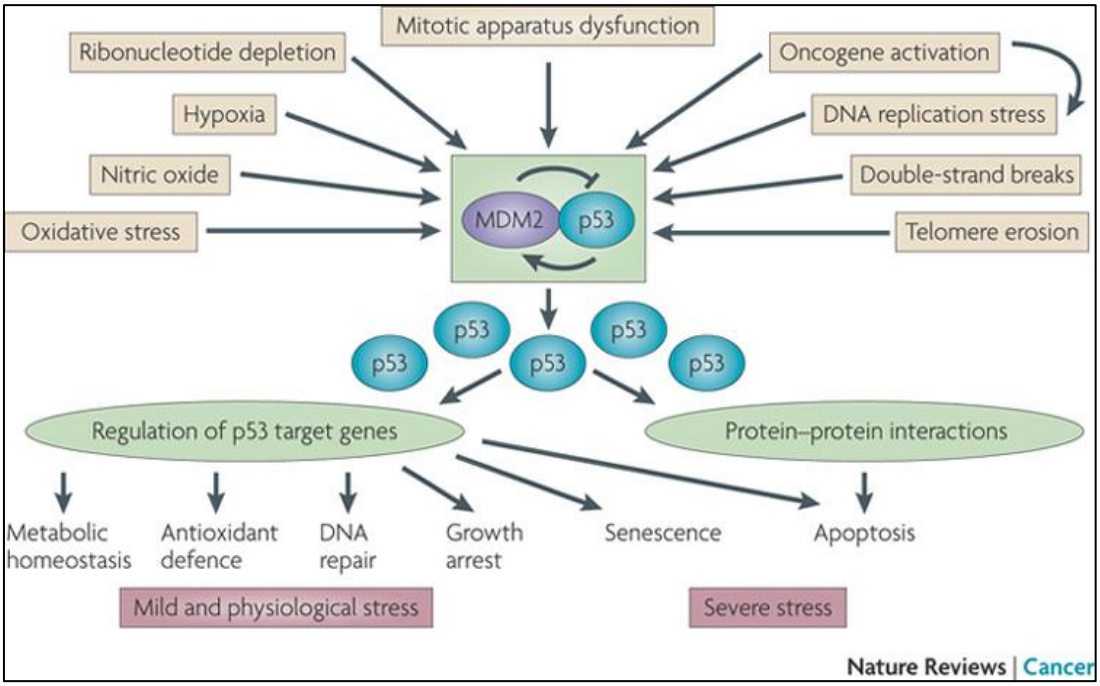


Figure 1.7. Different physiological stresses provoke the release of p53 from MDM2 inhibitory effect, thus increasing the concentration of p53. Once increased, the protein activates several different repairing mechanisms in order to correct the damage or, ultimately, cause irreversible cell arrest (senescence) or death (apoptosis). From (Levine and Oren, 2009).

c. Resisting cell death

Apoptosis, or programmed cell death, is a process that enables tissues to control the number of cells in a particular site and to remove cells termed to be unnecessary or defective (Elmore, 2007). The process is extremely important both during embryonic development and adult life. Cell proliferation and cell death, despite being opposite cellular fates, are remarkably interdependent and linked to each other by several molecular players (Lowe *et al*., 2004). In general, there are two different programmes which trigger apoptosis: the “intrinsic” or stress/mitochondrial pathway and the “extrinsic” or death receptor pathway (figure 1.8). The former activates

pro-apoptotic effectors present in the mitochondrial inter-membrane space. Mitochondrial membrane permeability is controlled by the balance between pro-apoptotic and anti-apoptotic factors. The Bcl-2 family of proteins is responsible for this equilibrium (Adams and Cory, 2007). The Bcl-2 protein was the first apoptotic regulator to be identified and was initially found in follicular lymphomas. Since its discovery, Bcl-2 hyper-expression has been demonstrated in a number of tumours and this expression is associated with resistance to chemotherapy-induced apoptosis (Amundson *et al*., 2000). At least other 16 members of the family were later discovered (Czabotar *et al*., 2014) and function as “life/death switches” of cellular fate. In a simplistic view, apoptosis occurs when the pro-survival Bcl2/Bclx$_L$ buffer is breached, allowing the dominance of the pro-apoptotic Bax/Bak subfamily, which leads to permeabilisation of the mitochondrial membrane and release of apoptotic factors such as cytochrome c and cell-death adaptor Apaf-1. These effectors, in turn, trigger the activation of caspase-9 leading to the initiation of a downstream proteolytic cascade that also includes caspases-3, -6 and -7 (Lowe *et al*., 2004, Czabotar *et al*., 2014). Once fully activated, caspases cleave proteins important for cell and genome integrity resulting in cell death. Cancer cells develop mechanisms that disrupt this balance and increase cell resistance to apoptosis. Malignant cells frequently exhibit down-regulation of pro-apoptotic proteins (such as Bax, Bak, Bid, Bik, Noxa, and Puma), down-regulation of death receptors and up-regulation of the anti-apoptotic mediators Bcl-2 and Bcl-x$_L$ (Safa, 2016). Furthermore, signalling pathways involved in proliferation and apoptosis resistance are also generally hyper-activated. For example, up-regulated PI3K/AKT/mTOR signalling is observed in many cancers (Safa, 2016).

Cancer cells can also develop mechanisms of resistance against treatment-induced cell death. Up-regulation of proteins of the ATP-binding cassette (ABC) family such as multidrug resistance (MDR) and multidrug resistance-associated proteins (MRPs) can cause the efflux of cytotoxic agents from the cell, thus reducing DNA damage (Al-Dimassi *et al.*, 2014, Safa, 2016). Breast cancer resistance protein (BCRP), another member of

the ABC family, is also over-expressed in a range of tumours, resulting in resistance against a variety of anticancer agents (Al-Dimassi *et al.*, 2014).

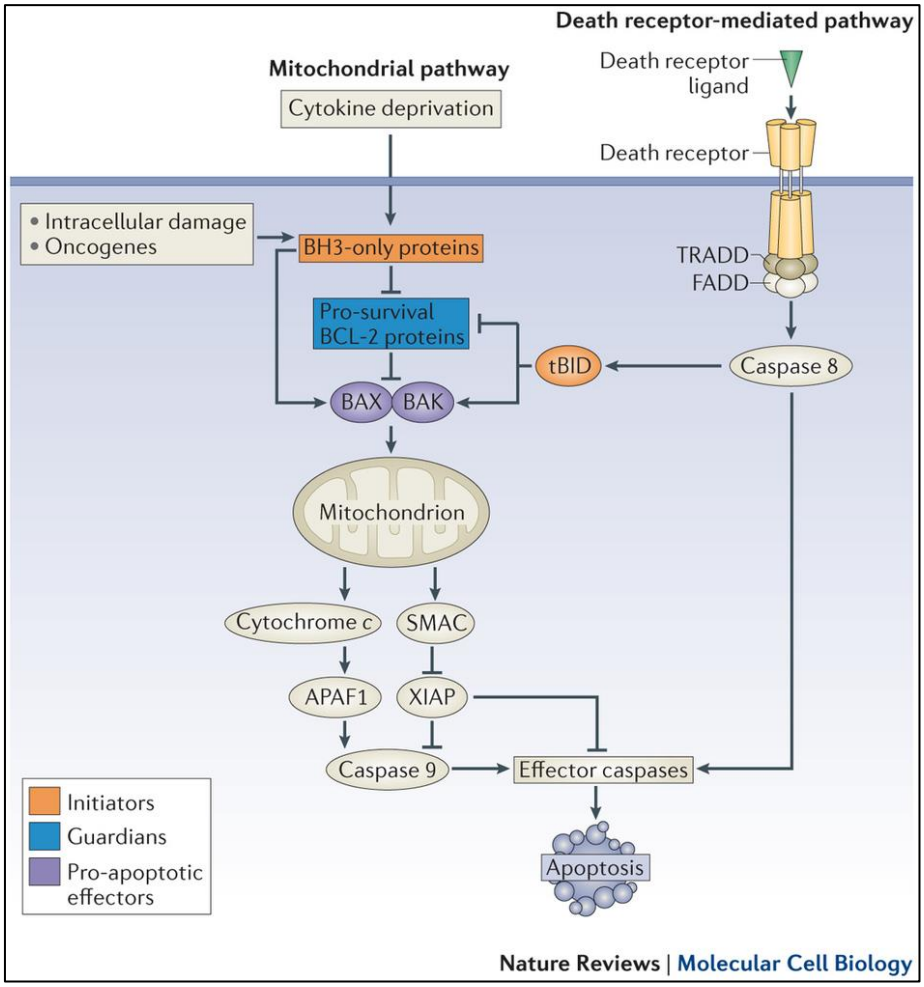


Figure 1.8. General overview of the "intrinsic" (also known as stress or mitochondrial pathway) and the "extrinsic" (or death receptor) apoptotic pathways. From (Czabotar *et al*., 2014).

On the other hand, the "extrinsic" pathway is activated through the binding of surface "death receptors" such as Fas/CD95, TNFR (tumour necrosis factor receptor) and DF-5 with their cognate ligands FasL, TNF-α and TRAIL (Lowe *et al*., 2004). When activated, these receptors form the "death-inducing signalling complex" (DISC) which causes the activation of apical caspase-8 (and caspase-10 in humans). These caspases can either trigger directly the downstream caspase cascade or recruit the mitochondrial pathway.

As described earlier, p53 has a pivotal role in the regulation of apoptosis. It performs this function both by acting as a transcription factor (inhibiting or stimulating the transcription of specific genes) and by acting directly on various elements of the regulatory machinery of the cell cycle (Levine and Oren, 2009). Several pro-apoptotic members of the Bcl-2 family of genes (including *puma, noxa, bid* and *bax*) are transcriptional targets for p53, as are some components of the death-receptor signalling complex (Fridman and Lowe, 2003). P53 can also facilitate cytochrome c release from the mitochondrial membrane. *TP53* mutations are some of the most common genetic alterations in cancer (Levine and Oren, 2009), and they severely compromise stress-induced and DNA damage-induced apoptosis.

d. Enabling replicative immortality

Besides the surveillance executed both by RB and p53, normal cells have an additional mechanism of proliferation control: a limitation in their replicative capacity. When in culture, normal cells are able to only replicate a certain number of times, after which they undergo senescence, a permanent non-replicative state and, then, apoptosis (Cristofalo and Pignolo, 1993, Hayflick and Moorhead, 1961). The main mechanism by which cell replication is limited is the presence of telomeres and the decreased (or absent) activity of the enzyme telomerase. Telomeres are the ends of linear chromosomes consisting in hexanucleotide tandem repeats of DNA associated with proteins (Blackburn, 1991). This repetitive terminal segment is responsible for protecting chromosomes against unions between the ends of non-homologous chromosomes. They allow the end of the linear chromosome to be replicated completely without loss of important coding bases at the 5’ end of each DNA strand (Blackburn, 1991, Shay and Wright, 2011). However, due to the unidirectional replication of DNA, the length of the telomere is shortened during each cell division. Therefore, after a certain number of cell cycles, the telomere becomes too short to guarantee chromosomal stability and DNA-damage signalling is activated, leading to cellular senescence (Blackburn, 1991). Thus, the presence of a telomere and its shortening after

each cell cycle limits the number of cell divisions and is a powerful protective measure against cancer development. Telomeres can be maintained by an enzyme called telomerase whose function is to specifically replicate the terminal portion of the telomere. When fully expressed, this enzyme is able to maintain telomere length and to prevent the occurrence of telomere shortening (Blackburn, 1991). Cells that physiologically sustain a high rate of proliferation such as embryonic, germline, hematopoietic and stem cells must undergo a disproportionately large number of cell divisions compared to most other cell types. In order to achieve that, these cells have active telomerase. However, the majority of adult human tissues lack any telomerase function (Collins and Mitchell, 2002). Consequently, to acquire unlimited replicative capacity, cancer cells must find ways to override telomere shortening. The preferential way by which cancers solve this problem is through the production of telomerase itself (Shay and Wright, 2011). About 80 to 90% of all tumours exhibit telomerase activity. How the remaining cases maintain their telomeres or replicative activity remains poorly understood (Günes and Rudolph, 2013).

e. **Inducing angiogenesis**

Oxygen and nutrients are fundamental for all types of tissues. Cancers are even more dependent on an adequate blood supply in order to sustain their high proliferative rates. It has been estimated that a solid tumour could grow only to a critical size of 1-2 mm (or about 10^6 cells) without the formation of new blood vessels (Carmeliet, 2005).

Angiogenesis is the synthesis of new blood vessels from pre-existing vessels, stimulated by the release of vascular growth factors e.g. vascular endothelial growth factor (VEGF) (Carmeliet, 2005, Prager and Poettler, 2012). VEGF, also named VEGF-A, is a member of the VEGF/Platelet derived growth factor (PDGF) family of structurally related mitogens. Its expression is up-regulated by a variety of growth factors including PDGF, fibroblast growth factor (FGF), epidermal growth factor (EGF), tumour necrosis factor (TNF), transforming growth factor beta (TGF-β) and

interleukin-1 (Carmeliet, 2005). Another important inducer of VEGF expression is hypoxia. VEGF mRNA expression is increased 10- to 50-fold in response to lowering the oxygen level from 21% to 0-3% (Shweiki *et al*., 1992). Therefore, as tumours grow, a multitude of growth factors operate in conjunction with the hypoxic state to generate large amounts of VEGF. Besides inducing endothelial cell proliferation and blood vessel growth, VEGF also promotes inhibition of apoptosis, increase in vascular permeability, chemotaxis and production of matrix proteases (Carmeliet, 2005).

Apart from the VEGF pathway, there are several less understood alternative angiogenic pathways such as the FGF, Notch Delta-like ligand 4 (DLL4) and PDGF pathways (Ferrara, 2010). A recent study showed that Norrin, a non-Wnt ligand capable of activating the canonical Wnt pathway, is produced by colorectal cancer cells and directly regulates endothelial proliferation (Planutis *et al*., 2014). The importance of these alternative angiogenic mechanisms is still to be fully understood.

f. **Activating invasion and metastasis**

Tumours are organised structures resembling organs, with different components and compartments. Apart from malignant cells, the tumour microenvironment encompasses the extracellular matrix (ECM), fibroblasts, mesenchymal stem cells, endothelial cells, immune cells and a network of cytokines and growth factors (Ye *et al*., 2014). The interaction between cancer cells and the microenvironment is fundamental for cancer initiation, invasion and metastasis (Talmadge and Fidler, 2010, Barkan *et al*., 2010, Friedl and Alexander, 2011, Bhowmick *et al*., 2004). The ECM is in immediate contact with the tumour cells and functions as a source of growth, survival, motility and angiogenic factors (Barkan *et al*., 2010). Fibroblasts are responsible for the synthesis, deposition and remodelling of the ECM and produce growth factors which stimulate tumour cells in a paracrine manner (Bhowmick *et al*., 2004).

The microenvironment is even more important for the development of metastasis - the spread and growth of tumour cells in distant organs. In 1889, the English surgeon Stephen Paget suggested that metastasis did not occur simply due to the arrest of tumour-cell emboli in the vasculature (which was the prevalent hypothesis at that time). Instead, he proposed the “seed and soil” hypothesis, according to which circulating tumour cells (“seeds”) should settle in a suitable tissue or organ (“soil”), establish a favourable interaction with the local environment and grow to form new tumours (Talmadge and Fidler, 2010). This explains why different types of tumours have different patterns of metastasis and different preferences for specific organs. Modern research has supported this concept and the interaction between tumour and environment in the development of metastasis has gained more importance (Langley and Fidler, 2011). Figure 1.9 shows an example of the multistage process of metastasis development.

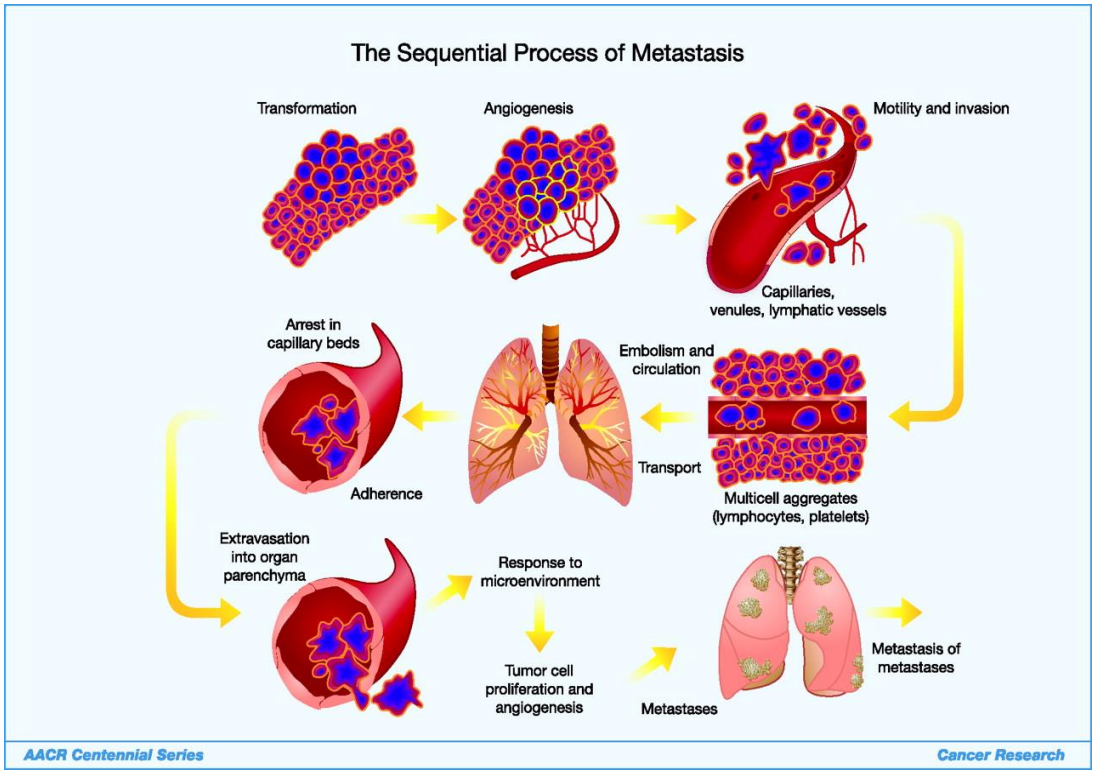


Figure 1.9. The multistage process of metastasis development. From (Talmadge and Fidler, 2010).

Another important aspect of metastasis is the heterogeneity of the tumour cell population. Different cancer cells, even within the same tumour, have different invasive and metastatic potentials (Talmadge and Fidler, 2010). Only a small fraction of the malignant cells are able to cause metastasis (Fidler, 1970) and this is probably due to the acquisition of genetic alterations which allow them to interact with the microenvironment in distant organs in a favourable way.

g. Genome instability and mutation

The occurrence of the constellation of genetic disturbances necessary for cancer formation is largely dependent on a succession of genomic alterations. Normal cells have an extraordinary ability to maintain genome stability. In order to overrule this protective mechanism, pre-cancerous cells develop defects affecting various components of the DNA-maintenance machinery, often referred to as the "caretakers" of the genome (Kinzler and Vogelstein, 1997).

Chromosomal instability (CIN) is one of the most consistent features of malignancy (Duesberg and Li, 2003). It is defined as a high rate of alteration in chromosomal structure and number. Besides CIN, other frequent forms of genomic instability are microsatellite instability – expansion or contractions of the number of oligonucleotide repeats in microsatellite sequences, and mutations in DNA-mismatch repair genes (Negrini *et al*., 2010). These alterations allow the pre-malignant cell to acquire cancerous hallmarks.

h. Tumour-promoting inflammation

It has long been known that tumours are infiltrated by distinct cells of the immune system. Historically, this immune infiltrate has been considered a defence mechanism by the host attempting to eradicate the tumour and its presence was viewed as a favourable prognostic factor (Pagès *et al*., 2010). However, chronic inflammation has also been observed prior to tumour

development in many scenarios. Clear examples of tumour-promoting inflammation are seen in gastric, colorectal and head and neck cancers, among others (Coussens and Werb, 2002). Even premalignant adenomatous lesions exhibit immune cell infiltration and dysregulation in several inflammatory cytokine genes (McLean *et al.*, 2011). Additionally, the use of aspirin, an anti-inflammatory drug, has been shown to consistently reduce the development of colorectal adenomas and current data also suggest a possible suppressive effect on CRC development (Drew *et al*., 2016). Taken together, these findings show that inflammation is sometimes an essential component of tumour development (Bondar and Medzhitov, 2013).

The mechanisms leading to tumour-promoting inflammation are not completely understood. However, several data show that the actions of cytokines, especially TNF-α, TGF-β and interleukins -1, -6, and -12 play an important role in the conversion from immune surveillance to tumour promotion (Mumm and Oft, 2008, Grivennikov and Karin, 2011). Furthermore, interleukin-10, a potent anti-inflammatory factor, appears to operate as a master switch in this transition by regulating other cytokines (Oft, 2014).

i. Reprogramming energy metabolism

It has recently been proposed that alteration in energy metabolism is also a “hallmark” of cancer (Hanahan and Weinberg, 2011). However, the first observations that cancer cells change their metabolic profile to counterbalance the shortage of oxygen and glucose (typically occurring in tumour growths) were provided by Otto Warburg many decades ago (Warburg *et al*., 1927). Since then, Warburg and other researchers have demonstrated that cancer tissues shift their metabolism from normal oxygen-consuming cell respiration (or mitochondrial oxidative phosphorylation) to oxygen-independent glycolysis, what is called the “Warburg-effect” (Hanahan and Weinberg, 2011, Hsu and Sabatini, 2008). During this process, cancer cells stop using oxygen in glucose metabolism and rely mostly on glycolysis to generate energy. Although advantageous in an oxygen-deprived

environment, glycolysis has an 18-fold lower efficiency in terms of energy generation compared to oxidative phosphorylation. Once driven by the "Warburg-effect", cancer cells maintain the preference for glycolysis even in the presence of oxygen. However, tumours are often also glucose-deprived due to the inadequate blood flow. To compensate for this, malignant cells up-regulate glucose transporters, notably GLUT1, which increases glucose transport to the cytoplasm (Hsu and Sabatini, 2008).

Lactic acid is the direct by-product of anaerobic glucose metabolism via glycolysis and is a prominent feature of tumour microenvironments (Denko, 2008). A recent study has tested the hypothesis that lactic acidosis could reduce glucose-dependence in cancer cells (Wu *et al*., 2012a). The authors exposed glucose-deprived 4T1 cell lines to media containing hydrochloric acid (acidosis), lactic acid (lactic acidosis), sodium lactate (lactosis) and control. They showed that the presence of lactic acidosis extended remarkably cellular survival compared to the other groups and that viable cells were still present even after 65 days of glucose deprivation. These findings may explain why tumours are more resistant to glucose starvation than normal tissues.

j. **Evading immune destruction**.

As mentioned earlier, immune defence has long been thought to play a major role in cancer control. Several experimental models (immunocompromised animals) and immunodeficient patients exhibit higher rates of cancer development (Buell *et al*., 2005, Swann and Smyth, 2007), further supporting this concept. Tumour infiltration by immune cells is associated with improved prognosis in some types of cancer (Pagès *et al*., 2010). Therefore, cancer cells must be able to circumvent immune surveillance and escape immune-mediated destruction (Swann and Smyth, 2007).

Initially, the prevalent hypothesis was that the immune system would exert a continuous surveillance against tumour formation. More recently, the

concept of immune editing has been widely accepted (Schreiber *et al*., 2011, Dunn *et al*., 2002). According to this model, the interaction between immune and malignant cells occurs in a three-step process, often referred to as the "three Es": elimination, equilibrium and escape. In the first phase, the immune system destroys the entirety or a portion of the tumour cell population. If some of the tumour cells remain after this initial attack, these malignant cells enter in a dynamic equilibrium with the immune cells, where lymphocytes exert a selective pressure that is effective enough to contain, but not fully destroy the many genetically unstable and mutating malignant cells. During this period, the remnant cancer cells develop additional mutations either reducing their immunogenicity or inhibiting the action of immune cells. After this, some malignant clones are able to avoid immune surveillance indefinitely leading to tumour growth and dissemination (Dunn *et al*., 2004).

1.3.2. **CRC-specific carcinogenesis**

Cancer is a common term used to describe more than a hundred different pathologies (Weinberg, 2007). Although most of them share the hallmarks described earlier, each cancer has a particular set of features. Since colorectal cancer is the focus of this research, an overview of the alterations most commonly present in CRC is described below.

a. The adenoma-carcinoma sequence

CRC provides an excellent system for the study of carcinogenesis in humans. It is one of the few cancer types in which it is possible to obtain biological samples from very early lesions to advanced tumours. It has long been recognised that the majority of colorectal tumours arise from benign polyps (Stryker *et al*., 1987, Muto *et al*., 1975). Several clinical-epidemiological data support this concept. For example, the prevalence of adenomatous polyps peaks 5 years earlier than the incidence of CRC (Muto *et al*., 1975) and the incidence of both lesions (polyps and cancer) shows a

strong correlation in different geographical regions (Clark *et al*., 1985). In addition, follow-up studies conducted prior to the introduction of endoscopic polypectomy showed that adenomas can undergo regression, growth or become malignant (Stryker *et al*., 1987, Leslie *et al*., 2002). Most colorectal polyps are hyperplastic, and this histological subtype does not seem to be associated with CRC. Conversely, the adenomatous polyp, or adenoma (a dysplastic epithelial lesion), is probably the most important precursor lesion (Jass, 2007). The estimated prevalence of adenomas in the United States population is 25% by age 50 and 50% by age 70 (Rex *et al*., 1993).

Around 25 years ago, Vogelstein and Fearon proposed a model for colorectal carcinogenesis which suggested that CRC occurs as a result of the accumulation of inactivating mutations in tumour suppressor genes and activating mutations in oncogenes (Vogelstein *et al*., 1988, Fearon and Vogelstein, 1990). The most common alterations found in CRC, at that time, were activating mutations in the oncogene *KRAS* and inactivations and/or deletions in the chromosomal regions 5q (later found to be the site of the *APC* gene), 17p (where the *TP53* gene is located) and 18q (*SMAD4* gene site) (Fearon and Vogelstein, 1990). Later, it became clear that inactivation of *APC* and activation of *KRAS* were early events in CRC carcinogenesis. On the other hand, mutations of *TP53* and various other genes accumulated later during tumour development (Ahnen, 2011). On average, there are around 80 mutations in the genome of the malignant cells in a typical case of CRC (Wood *et al*., 2007). Notably, only about 15 of these mutations contribute actively to the malignant phenotype and are referred to as “driver” mutations. Figure 1.10, shows the proposed progression from normal epithelium to adenoma to carcinoma and the most commonly involved genetic alterations.

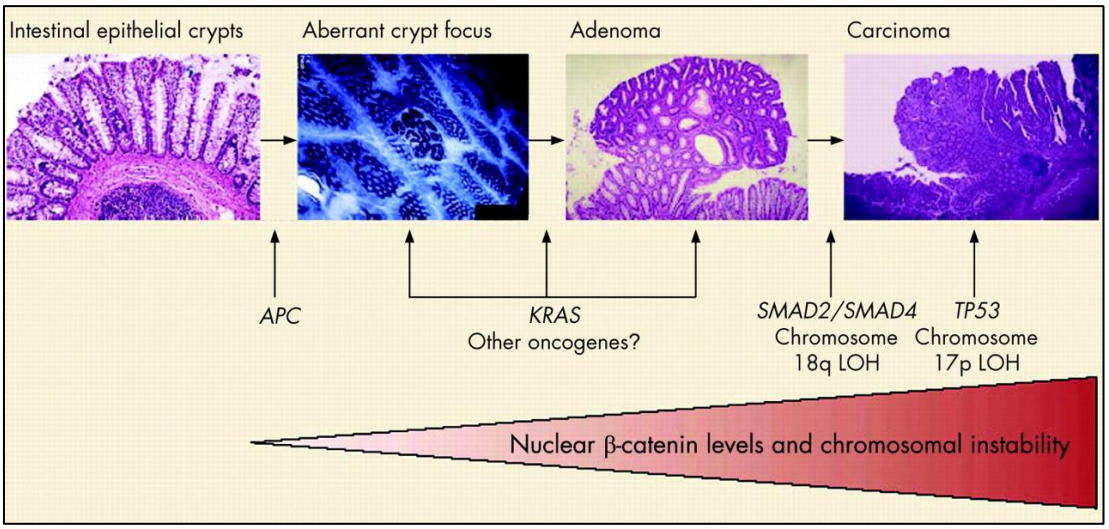


Figure 1.10. Genetic model of colorectal carcinogenesis. The illustration highlights the main genetic alterations which drive CRC development. From (Leedham and Wright, 2008).

b. Wnt signalling pathway

Among several signalling pathways involved in carcinogenesis, the Wnt pathway plays a particularly important role in CRC development. This signalling cascade is normally involved in controlling embryonic development and tissue homeostasis in adults (Holland *et al*., 2013, Clevers and Nusse, 2012). It is named after the fusion of two gene names: *wingless* (*wg*) and its vertebrate homolog *integrase 1* (*int-1*) (Baarsma *et al*., 2013). The mammalian genome encodes 19 Wnt proteins and 10 Frizzled (FDZ) membrane receptors (van Amerongen *et al*., 2008). According to the results of the Cancer Genome Atlas Network, the Wnt signalling pathway is altered in 93% of CRC cases (Cancer Genome Atlas Network, 2012). There are two theoretical mechanisms for Wnt activation: the canonical and the non-canonical signalling pathways. As figure 1.11 illustrates, the canonical (or β-catenin – dependent) pathway depends on the binding of Wnt ligands to FZD receptors on the cell membrane. In the absence of Wnt ligands, FZD is inactive and, as a consequence, the protein Dishevelled (DVL) is also non-phosphorylated and inactive (Baarsma *et al*., 2013). This inactivation allows a group of proteins (APC, Axin, CK-1 and GSK-3) to work as a β-catenin-targeted “destruction complex”. This complex phosphorylates β-catenin,

which is then destroyed by the proteasome system (Baarsma *et al*., 2013). By contrast, in the presence of Wnt ligands, FDZ becomes active, associates with LPR5/6 – another membrane receptor, and phosphorylates DVL, which results in disengagement and disruption of the "destruction complex". As a result, β-catenin is no longer phosphorylated and destroyed, and accumulates in the cytoplasm. Then it translocates to the nucleus and, in conjunction with the TCF/LEF family of transcription factors, activates its target genes (Baarsma *et al*., 2013, Clevers and Nusse, 2012). The precise mechanism by which this happens is still poorly understood. Some classical examples of Wnt/β-catenin target genes are *c-Myc* (He *et al*., 1998) and *CCND1* (Tetsu and McCormick, 1999). However, several other genes have been added to the group of targets and an updated list can be found at the Nusse Laboratory website (http://web.stanford.edu/group/nusselab/cgi-in/wnt/target_genes). Overall, 80% of CRCs have defects that lead to either APC inactivation (majority of cases) or β-catenin over-expression (minority of cases), both resulting in up-regulation of the Wnt pathway (Cancer Genome Atlas Network, 2012).

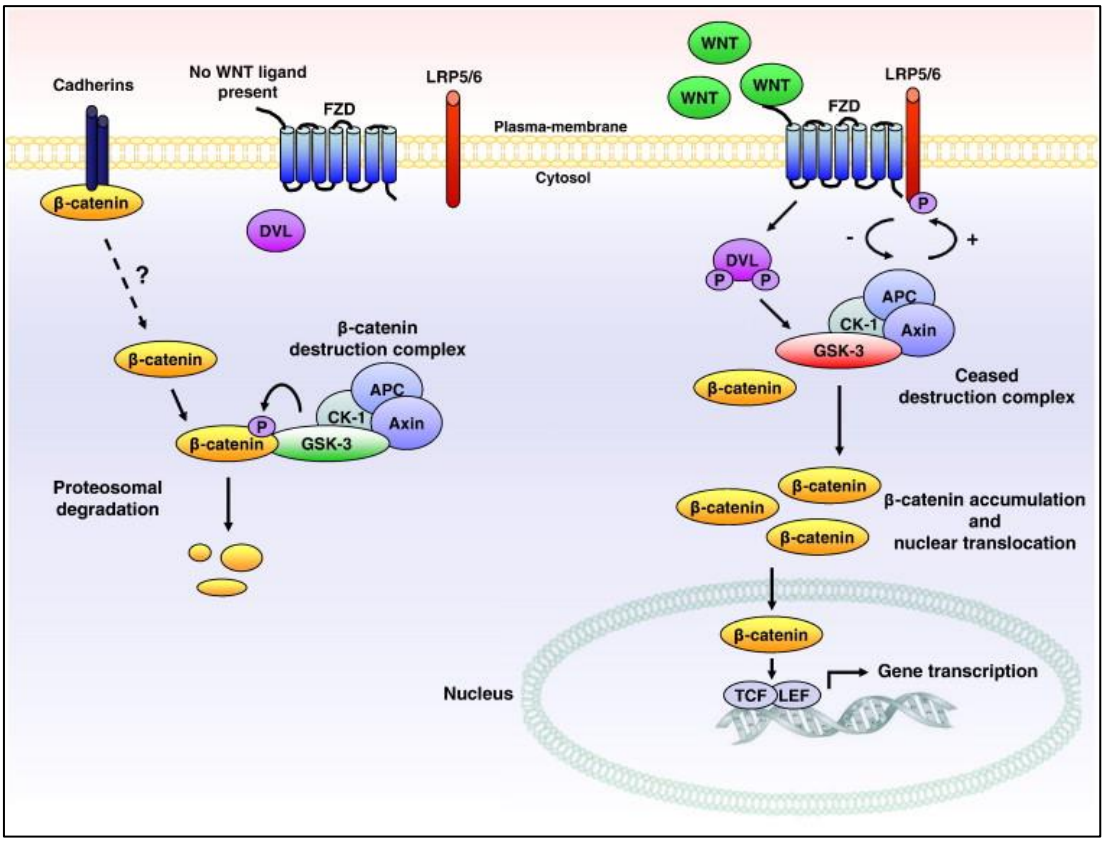


Figure 1.11. The canonical Wnt signalling pathway. The illustration shows the actions of Wnt pathway elements in the absence (left) or presence (right) of Wnt ligands. From (Baarsma *et al*., 2013).

The other postulated mechanism for Wnt activation is the non-canonical pathway. It is defined as Wnt- or FDZ-initiated signalling that is independent of β-catenin transcriptional function (Semenov *et al*., 2007). Several different and less characterised functions have been attributed to the non-canonical Wnt pathway including planar cell polarity, calcium influx, bone metabolism and control of atypical protein kinase C pathways, to cite a few (Semenov *et al*., 2007, van Amerongen, 2012). Wnt3a (a prototypical canonical Wnt ligand) and Wnt5a (a prototypical non-canonical ligand) are thought to trigger Ser/Thr phosphorylation on LRP6 and Ror2 (a non-canonical membrane receptor) respectively, using common intracellular components (van Amerongen *et al*., 2008, Grumolato *et al*., 2010). After activation of the receptors, intracellular signalling cascades are triggered targeting their respective effectors and genes.

However, there has been a recent change in the way these two apparently distinct pathways are viewed. It has been shown that the type of receptor displayed on the cell surface rather than the type of ligand is the most important factor in defining canonical or non-canonical activation (van Amerongen *et al*., 2008, van Amerongen, 2012). In addition, it is now known that Wnt ligands from either "class" can elicit both β-catenin-dependent and independent responses depending on the receptor profile of the cell (van Amerongen *et al*., 2008). Therefore, our knowledge about the Wnt pathways is evolving and it is expected that in the future we will have a better understanding of their roles in physiological and pathological conditions, thus allowing the development of Wnt-targeted tools for diagnosis and treatment (Anastas and Moon, 2013).

c. Adenomatous Polyposis Coli (*APC*) in hereditary and sporadic CRC

Since the 1920s, a hereditary form of CRC has been recognised in which the occurrence of the tumour was preceded by the development of hundreds to thousands of intestinal adenomatous polyps (Plawski *et al*., 2013). The disease exhibited an autosomal dominant inherited pattern and had a nearly complete penetrance for CRC by the age of 50 years. Familial Adenomatous Polyposis, or FAP, was frequently associated with deletions of the chromosomal region 5q21. In the early 1990s, germline mutations at that site (later named Adenomatous Polyposis Coli gene - *APC*) were found to be the cause of this hereditary colorectal cancer syndrome (Kinzler *et al*., 1991, Nishisho *et al*., 1991), providing the first direct link between the Wnt signalling pathway and human disease. Several types of *APC* mutation occur in FAP individuals. The most common are small deletions leading to changes in the reading frame and the creation of premature termination codons (Plawski *et al*., 2013, Kerr *et al*., 2013). Other forms are large deletions, splice site mutations, small deletions plus insertions, large insertions, complex rearrangements and mutations in regulatory sequences (Plawski *et al*., 2013). Apart from the classical FAP syndrome, other forms of

“attenuated” FAP are characterised by fewer polyps (usually less than 100) and a lower risk of malignant transformation. These have been associated with mutations at the 5’ end of *APC* gene and alternative splicing (Plawski *et al*., 2013).

After the unequivocal demonstration that germline *APC* mutations are the cornerstone of hereditary FAP, *APC* somatic mutations were also demonstrated in the vast majority of sporadic CRC cases, establishing this genetic defect as the earliest and most common mutation in cancers of the colon and rectum (Cancer Genome Atlas Network, 2012).

d. β-catenin as a prognostic factor

The expression of β-catenin in CRC tissues has been evaluated in several studies (Chen *et al*., 2013). The majority have analysed the associations between the expression of β-catenin and clinicopathologic variables such as stage, differentiation and prognosis. The effect of β-catenin accumulation on the survival of patients is controversial. A meta-analysis addressing this specific question included 18 studies (analysing β-catenin expression by immunohistochemistry) and found heterogeneous results (Chen *et al*., 2013). Nonetheless, it suggested that the nuclear accumulation of β-catenin, but not the cytoplasmic expression of this protein, is associated with advanced tumour stage and worse prognosis. Reinforcing the contradictory findings, a recent large study not included in that meta-analysis showed that loss of membranous and cytoplasmic β-catenin was associated with shortened survival, but its nuclear expression was not associated with any significant outcome (Bruun *et al*., 2014). Therefore, despite the multitude of research conducted in this field, the definitive role of β-catenin expression in terms of CRC prognosis is far from elucidated.

e. *C-MYC* is an essential target for Wnt/β-catenin activation

MYC lies at the crossroad of many growth-promoting signalling pathways (Dang, 2012). The protein was first reported in association with the ***MY****elo****C****ytomatosis* viral oncogene in animal models and is now known to be one of the most amplified oncogenes in several different human cancers (Beroukhim *et al*., 2010). *MYC* mRNA is short-lived and affected by different microRNAs. The protein itself is post-translationally modified, ubiquitinated and degraded quickly, with a half-life of 15-20 minutes (Gregory and Hann, 2000). Mutations in the residues Thr-58 and Ser-62 (the targets for phosphorylation and degradation) are common in Burkitt lymphomas and cause stabilisation of the protein (Thomas and Tansey, 2011, Wang *et al*., 2011). One of the main mechanisms by which MYC works is as a transcription factor. A study mapping potential Myc binding sites throughout the genome found that up to 6,000 genes could be targets for this oncoprotein. Of these, 700 responded to MYC activation with alterations in their mRNA levels (Zeller *et al*., 2006). Other MYC functions involve non-transcriptional actions, regulation of microRNAs and recruitment of histone-acetylases (Dang, 2012).

Myc has been demonstrated to be an efficient oncogene in murine models and cell lines, and its knock-down in cancer cells is associated with reduced proliferation and increased apoptosis (Dang, 2012). The same anti-tumour effect was demonstrated by inhibiting endogenous Myc in animal models of lung and pancreatic cancer (Soucek *et al*., 2008, Sodir *et al*., 2011).

This oncogene is overexpressed in most CRC cases (Chan *et al*., 2008). However, the gene is rarely mutated in this disease (Cancer Genome Atlas Network, 2012), suggesting that an upstream factor is responsible for inducing *MYC* transcription. MYC is a known target for the Wnt pathway (He *et al*., 1998). Several experiments have shown that either APC inactivation or β-catenin activation result in overexpression of this gene in human and animal models (He *et al*., 1998, Sansom *et al*., 2004). Myc contributes to all the phenotypic changes observed in mouse intestinal mucosa after *Apc*

deletion such as alterations in cell proliferation, differentiation, migration and tissue architecture. Remarkably, co-deletion of both *Apc* and *Myc* results in a normal intestinal phenotype, thus providing strong evidence that Myc is not only important but rather essential for intestinal tumourigenesis (Sansom *et al*., 2007, Myant and Sansom, 2011).

Interestingly, despite the role of *MYC* in cancer development, its over-expression in CRC samples determined by Northern blot analysis was associated with good prognosis (Smith and Goh, 1996). Recent work using immunohistochemistry found similar results, and showed that patients whose tumours over-expressed nuclear MYC had better survival rates than those with low nuclear expression (Toon *et al*., 2014).

f. DNA mismatch repair defects – another important cause of sporadic and hereditary CRC

The DNA mismatch repair (MMR) system is one of the enzymatic mechanisms responsible for genome homeostasis (Sameer *et al*., 2014). It is involved in the identification and repair of specific types of errors in DNA replication. When this system fails, the rate of errors increases 100- to 1000-fold, particularly in so-called DNA microsatellites – short tandem repeats of between 1-5 DNA bases scattered throughout an individual's genome (Sameer *et al*., 2014).

The main genes associated with MMR in eukaryotes are *MutS* homologs (*MSHs)*, and *MutL* homologs (*MLHs)*. There are five highly conserved *MSHs* (*MHS2-MSH6*) and four *MLHs* (*MLH1, MLH3*, *post meiotic segregation 1* or *PMS1*, and *PMS2*) (Jun *et al*., 2006). If any of those MMR elements is inactive, DNA is no longer replicated with the necessary high fidelity, thus allowing oncogenic mutations to occur.

A defect in MMR is found in 15% of sporadic CRC cases and is the second most important mechanism responsible for colorectal carcinogenesis (Sameer *et al*., 2014). Most of these cases are caused by epigenetic silencing of MMR genes, *MLH1* being the most commonly affected (Li *et al*.,

2013b, Iacopetta *et al*., 2010). Due to the fact that microsatellites are parts of the DNA that are particularly sensitive to replication errors when any MMR mechanism is inactivated, it often results in microsatellite instability (MSI). MSI is defined as a change of any length caused by either insertions or deletions of repeating units in a microsatellite within a tumour compared to normal tissue (Sameer *et al*., 2014). MSI is categorised using a panel of five allocated microsatellite loci or markers known as the Bethesda panel. MSI-high frequency (or MSI-H) is defined by the presence of instability in 2 or more markers, or more than 30% of a larger panel; MSI-low frequency (or MSI-L) occurs when only one marker is affected (or 10-30% of a larger panel); and microsatellite stability (or MSS) is characterised by absence of instability in any marker (Boland *et al*., 1998). Therefore, the presence of MSI-H is a surrogate marker for a MMR defect in colorectal cancer.

Apart from the importance of MMR inactivation in sporadic CRC development, it has also great importance in the origin of a type of hereditary cancer syndrome. A familial predisposition for CRC in patients without intestinal polyposis (contrasting with FAP patients) was initially described in 1913 in the United States (Warthin, 1913). Many years later, Henry Lynch described two large families with similar characteristics: predisposition to colorectal and other cancers (endometrial and stomach), thus defining a condition termed *hereditary non-polyposis colorectal cancer* (HNPCC), also known as Lynch syndrome (Lynch *et al*., 1966, Lynch *et al*., 1999). It is now known that 90% of HNPCC cases are caused by autosomal-dominant mutations in either *MLH1* or *MSH2* (Sameer *et al*., 2014, Geiersbach and Samowitz, 2011). As depicted in figure 1.12, HNPCC is the most common form of hereditary CRC cancer followed by FAP.

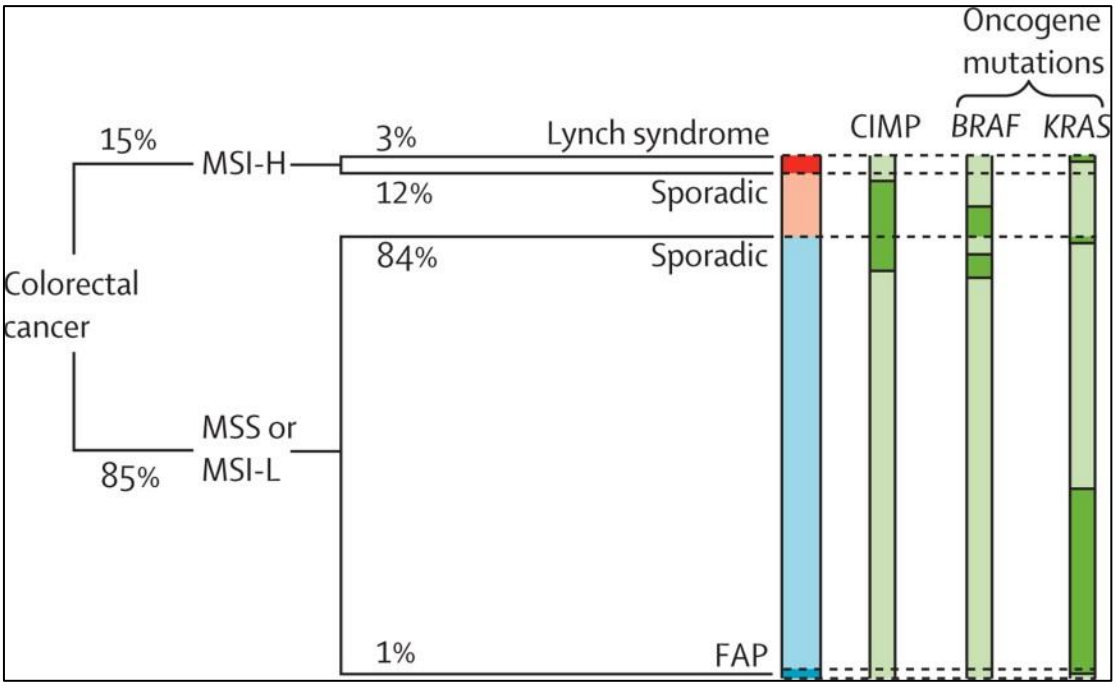


Figure 1.12. Molecular classification of CRC. The figure defines various categories of CRC according to the genetic origin (Lynch vs FAP vs sporadic) and the status of the MMR machinery (MSI-H vs MSI-L/MSS). Additionally, it shows the proportional contributions of CIMP and important oncogene mutations (*KRAS* and *BRAF*) for colorectal carcinogenesis. From (Brenner *et al*., 2014).

g. CpG island methylator phenotype (CIMP)

A third mechanism responsible for CRC development has recently been proposed and is now widely accepted: the epigenetic alteration of gene expression (Bae *et al*., 2013). Epigenetic modifications are changes in gene expression that are not associated with alterations in DNA sequence. Promoter methylation is one of the best known mechanisms of epigenetic silencing. DNA methylation occurs at the 5' cytosine moiety of CpG nucleotides. In humans, CpG nucleotides are aggregated (forming "CpG islands") in certain regions of gene promoters. It has been demonstrated that a concordant pattern of hypermethylation of multiple CpG island loci is present in a number of CRC cases. This is referred to as the *CpG island methylator phenotype*, or CIMP (Toyota *et al*., 1999). Following these results, researchers have confirmed the presence of CpG island hypermethylation in the promoter of several genes known to be associated with colorectal

carcinogenesis, such as *APC, MCC, MLH1, MGMT, SMAD4, MSH6, p16* and others (Sameer *et al*., 2014).

CIMP-positive tumours do not follow the classical adenoma-carcinoma sequence. The precursor lesion which gives rise to these cancers is the so-called sessile serrated polyp/adenoma (Yamane *et al*., 2014). These lesions are normally flat (sessile), frequently located in the proximal colon and resemble the architecture of hyperplastic polyps. Serrated carcinomas can originate from the precursor lesion and exhibit high rates of *BRAF* mutations and MSI (Bae *et al*., 2013). Besides this distinct mechanism, CIMP-positive tumours also exhibit different epidemiological and pathological features such as predominance in women, older age of occurrence, increased production of mucin, increased de-differentiation and, as in the precursor lesions, higher rates of MSI and *BRAF* mutation (Bae *et al*., 2013, Nazemalhosseini Mojarad *et al*., 2013).

h. Other important pathways and genes

There are many other important pathways and genes involved in colorectal carcinogenesis. Alterations in these elements are present in a sizable proportion of sporadic CRCs (usually developed later during multistep carcinogenesis) and are the cause of a small fraction of hereditary cancer cases (Fearon, 2011).

The Ras family of small guanosine nucleotide binding proteins function as molecular switches downstream of growth factors. These proteins participate in the regulation of many different intracellular signalling pathways such as the PI3K-Akt-mTOR and the Ras-Mek-Erk pathways (Malumbres and Barbacid, 2003), as illustrated in figure 1.13. The three members of the Ras family – *KRAS, HRAS* and *NRAS* – are common sites of somatic mutation in human cancers. *KRAS* mutations are present in 40% of CRCs while another small fraction of colorectal tumours have *NRAS* mutations (Fearon, 2011).

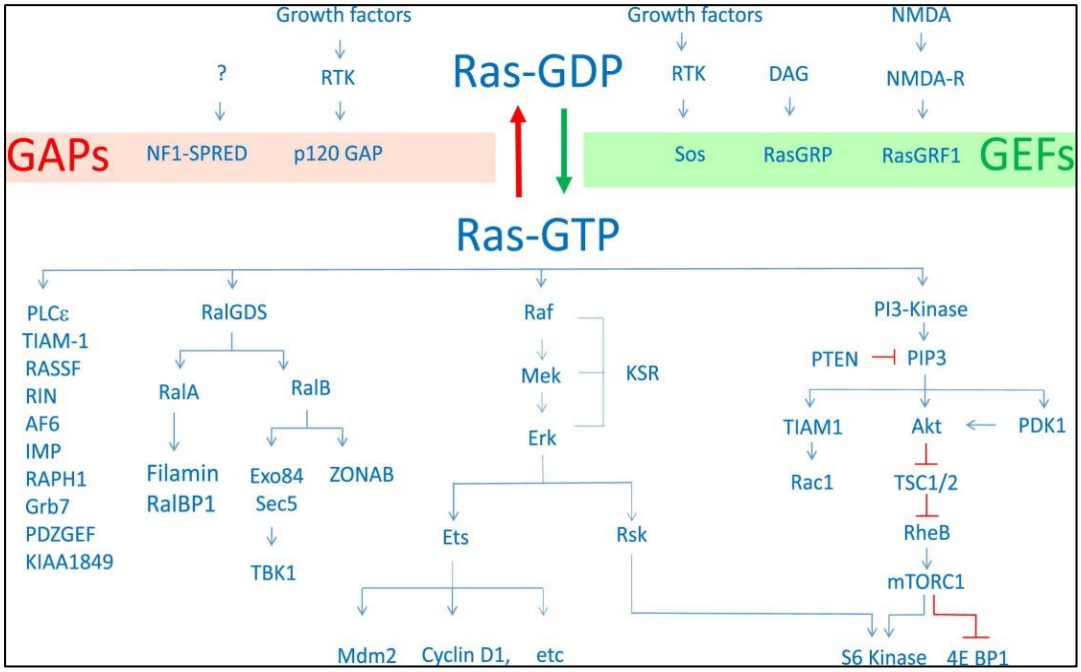


Figure 1.13. A simplified view of the Ras pathway. Under stimuli from different sources (especially growth factors), Ras protein changes from GDP- to GTP-status and activates several signalling pathways. GAP: GTPase activating protein. GEF: guanine nucleotide exchange factor. RTK: receptor tyrosine kinase. From (Stephen *et al*., 2014).

Although *RAS* mutations are common in CRC, mutations in other elements of the RAS pathway are less frequent. For example, mutation in the *EGFR* gene (an upstream regulator of Ras) occurs in only 5% of CRCs and *RAF* mutations are also present in the same proportion of cases (most being associated with CIMP, as mentioned earlier) (Fearon, 2011). *PIK3CA*, the gene responsible for the synthesis of PI3K, is mutated in approximately 15-25% of CRCs and *PTEN* mutations are present in 10% of sporadic cases (Fearon, 2011).

Loss-of-heterozygozity (LOH) is a key mechanism for inactivating a normal allele in heterozygous tumour suppressor genes, thus leaving only one mutated allele. Around 70% of CRCs exhibit LOH on chromosomal regions 17p and 18q (Fearon, 2011). Region 17p harbours the tumour suppressor gene *p53*, which is thought to be the main target of LOH because, in most cases, the remaining *TP53* allele carries a somatic mutation (Baker *et al*., 1990b). However, the same appears to be the case for

SMAD2 and *SMAD4* genes which are located at the 18q region (Fearon, 2011). These genes encode proteins that function downstream to the TGF-β receptor complex. Both the occurrence of SMAD mutations (5-15% of CRCs) and TGF-β receptor mutations (around 25% of CRCs) (Grady *et al*., 1999) result in inactivation of the TGF- β tumour suppressor pathway.

As is the case for most cancer types, with the progression of the carcinogenic process, many other mutations accumulate and any attempt to address this multitude of genetic alterations would be unsuccessful and only contribute to confusion. Only the most important and frequently involved genes/pathways have therefore been mentioned in order to provide a more general overview of colorectal carcinogenesis.

i. The consensus molecular subtypes of CRC

Several molecular classifications of CRC have been proposed. Given the broad interconnectivity between these categorisations, an international consortium was formed in order to assess the different systems and this group proposed a unified molecular classification of CRC (Guinney *et al.*, 2015). These consensus molecular subtypes (CMS) identify most of the genetic, molecular and cellular events related to CRC carcinogenesis and discussed above. Figure 1.14 shows the CMS, the proportion observed for each category and the main cellular and molecular events observed. Briefly, CMS1 tumours exhibit MSI and CIMP features. Immune cell infiltration is also commonly observed. CMS2 is marked by Wnt/MYC activation and somatic copy number alterations. In CMS3 tumours, metabolic dysregulation is the hallmark. CMS4 tumours exhibit features of epithelial-mesenchymal transformation (Guinney *et al.*, 2015). Classification of CRC based on molecular characteristics rather than morphology may improve cancer treatment by allowing the use of therapies targeting biologically relevant molecular events (Dienstmann *et al*., 2017).

CMS1 MSI immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
14%	37%	13%	23%
MSI, CIMP high, hypermutation	SCNA high	Mixed MSI status, SCNA low, CIMP low	SCNA high
BRAF mutations		*KRAS* mutations	
Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration, TGF-β activation, angiogenesis
Worse survival after relapse			Worse relapse-free and overall survival

Figure 1.14. The consensus molecular subtypes (CMS) of CRC. CIMP, CpG island methylator phenotype; MSI, microsatellite instability; SCNA, somatic copy number alterations. From (Guinney *et al.*, 2015).

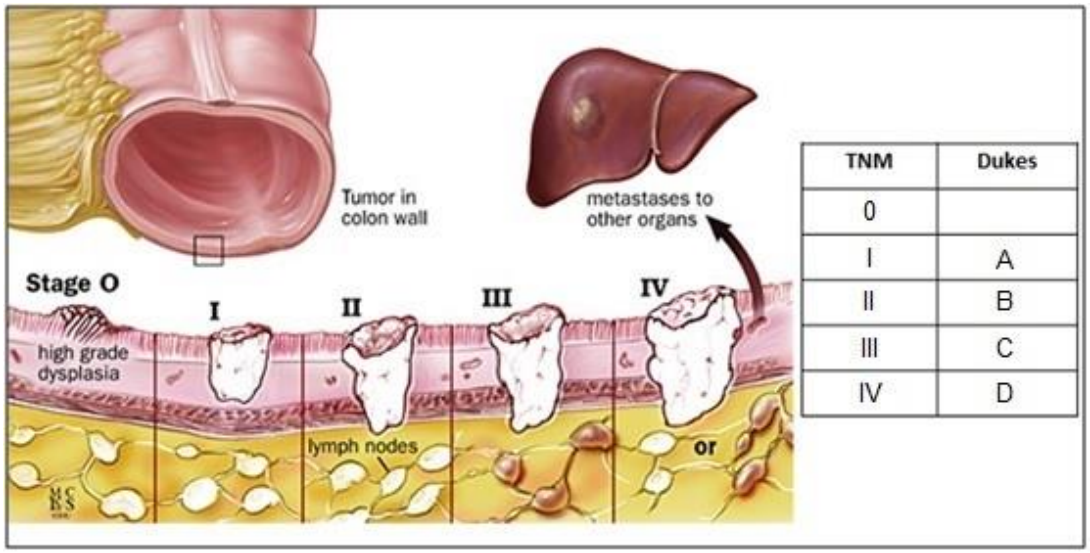
1.4. Clinical aspects of CRC

Colorectal tumours arise from premalignant lesions (polyps) in a process which takes up to 10 to 20 years to be completed (Kuntz *et al*., 2011). During most of this period, the disease is asymptomatic and can only be detected through active screening. Clinical manifestation usually occurs when the disease is sufficiently advanced to cause complications. The most common signs and symptoms associated with CRC are rectal bleeding, change in bowel habit (diarrhoea, constipation or difficult evacuation), weight loss, abdominal pain and anaemia due to iron deficiency (John *et al*., 2011).

When CRC is suspected from a patient´s history and physical examination (including digital rectal examination), imaging studies of the colon are recommended. Barium enema x-ray can provide an adequate evaluation of the colon. However, it is less sensitive for small lesions and requires endoscopic evaluation to confirm the nature of any alteration found (Rockey *et al*., 2005). Most clinicians therefore now favour colonoscopy as the first line test for CRC diagnosis. Colonoscopy permits the evaluation of the entire colon, resection of small premalignant lesions and biopsy of larger

polyps and tumours. The majority of intestinal cancers are adenocarcinomas or its variations (mucinous, signet-ring type, etc.) (Quirke *et al*., 2012). These lesions are histologically categorised according to the degree (or grade) of differentiation – a marker for tumour aggressiveness. Well-differentiated tumours maintain a relatively organised architecture resembling the normal colon whereas undifferentiated lesions exhibit higher mitotic rates, no tissue organisation and no similarities with the normal epithelium. Other features detailed on a pathology report are tumour size, depth of invasion, lymphovascular or lymph node invasion, resection margin involvement and the presence of tumour budding (Quirke *et al*., 2012).

Once a CRC diagnosis has been made, the next clinical step is the evaluation of disease extension, a process called staging. Physical examination and image analysis are performed in order to identify signs of tumour spread both locally and at distance. X-rays, computed tomography scans, magnetic resonance imaging and positron-emission tomography (PET) are all used for this purpose (Dewhurst *et al*., 2012). Two staging systems are mainly used for CRC classification: Dukes' staging and TNM (tumour-node-metastasis) staging. Figure 1.15 illustrates and compares these systems. Cancer staging provides fundamental information about the prognosis and guides the modalities of treatment that should be employed. Overall survival is closely related to stage at diagnosis as shown in figure 1.16.



TNM	Dukes
0	
I	A
II	B
III	C
IV	D

Figure 1.15. CRC staging systems. On the left, the figure shows the extension of disease in each stage. Stage 0 is defined as the presence of premalignant lesion (high grade dysplasia or carcinoma *in situ*). In stage I, an invasive cancer is limited to the intestinal mucosa or submucosa. In stage II, the tumour extends beyond the *muscularis propria* but is confined to the intestinal wall. In stage III, the tumour spreads to regional lymph nodes. Finally, stage IV is defined by the presence of metastasis in distant organs. On the right, a table comparing TNM and Dukes' staging is provided. Figure adapted from http://www.hopkinscoloncancercenter.org/

Dukes' stage at diagnosis	Percentage of cases	Five-year survival
A	8.7%	93.2%
B	24.2%	77.0%
C	23.6%	47.7%
D	9.2%	6.6%
Unknown	34.3%	35.4%

Figure 1.16. Distribution of stages at diagnosis and 5-year survival of patients with colorectal cancer. Figure from: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/ survival

The only curative treatment for CRC patients is the surgical removal of the tumour. However, many patients with apparently limited disease who undergo resection will ultimately relapse and die from CRC. For that reason, other treatment modalities are frequently used to try to improve the cure rate.

Chemotherapy is usually administered after surgery for patients with late stage disease (i.e. stages III and IV) and has resulted in improvements in cure rates and overall survival (Tsai *et al*., 2011, Chen *et al*., 2010b). Radiation therapy is also frequently used for patients with rectal tumours before or after surgery to improve local disease control. The use of preoperative radiotherapy also increases the chance of anal sphincter preservation (Glimelius, 2013). The majority of patients with metastatic disease are incurable and the main goal of treatment in these cases is to alleviate symptoms and prolong survival. The use of palliative chemotherapy has resulted in significant prolongation of survival compared to supportive care alone (Chen *et al*., 2010b, Funaioli *et al*., 2008).

The knowledge gained from the study of the molecular biology of CRC has translated into new forms of therapy for patients with advanced disease. Drugs targeting specific molecular players or pathways involved in tumour development are now being routinely used in the treatment of patients with CRC. For example, the angiogenic pathway has been targeted by using inhibitors of VEGF (Prenen *et al*., 2013). Anti-VEGF agents currently available for clinical use include Bevacizumab, Aflibercept and Regorafenib. Additionally, the EGF pathway is another successful target in CRC treatment. Cetuximab and Panitumumab are monoclonal antibodies which bind to the EGF receptor causing its inactivation (Prenen *et al*., 2013). Although the inhibition of VEGF or EGF signalling results in clinical benefits, the mean survival of patients with metastatic disease rarely exceeds 24 months even with the use of these agents (Dattatreya, 2013). Efforts to find biomarkers to predict a higher chance of benefit from these new therapies are beginning to produce good results. An illustrative example is the identification of *KRAS* mutation as a negative predictor of response to EGFR-targeted drugs, thus allowing the selection of only *KRAS*-wild type patients to receive these treatments (De Roock *et al*., 2010). This selection has resulted in improvements in response rate and overall survival in this group of patients (Prenen *et al*., 2013). These achievements highlight the importance of identifying more predictive biomarkers to guide cancer treatment.

1.5. CRC screening

As explained earlier, colorectal tumours usually develop following a well-defined stepwise process from normality to cancer. This transformation is thought to occur over many years and thus offers an excellent window of opportunity for tumour detection (Kuntz *et al*., 2011, Kuipers *et al*., 2013). Therefore, CRC is a suitable candidate for population-based screening strategies. Various screening methods have been developed and tested over the last few decades (Kuipers *et al*., 2013). Some of these methods have now been validated and have been adopted in several regions and countries (Benson *et al*., 2008). However, there are many drawbacks associated with the current screening tools and research still continues with the aim of improving the accuracy and compliance with CRC screening (Kuipers *et al*., 2013). This section provides an overview of the currently available screening methods and addresses some of the new strategies in this field.

1.5.1. **Current methods of CRC screening**

Current CRC screening tests are broadly grouped as early detection tools or cancer-prevention tools. Early detection tools include faecal occult blood tests (FOBTs) and faecal immunological tests (FITs). These are cheap, non-invasive methods which detect microscopic amounts of blood in the stool (Kuipers *et al*., 2013). When positive, the individual must undergo an additional test to confirm the presence of a lesion. This is usually done by endoscopic examination of the rectum and sigmoid (flexible sigmoidoscopy, FS) or the entire large bowel (colonoscopy). When used as the primary screening tool, FS and colonoscopy are termed cancer-prevention tools due to their ability to both diagnose and treat precursor lesions (Kuipers *et al*., 2013).

Guaiac faecal occult blood test (gFOBT)

A gFOBT card uses paper impregnated with guaiac, a phenolic compound present in the resin of a South American hardwood tree (*Guaiacum officinale*), onto which faeces is applied. In the presence of haem, a component of haemoglobin, oxygen released from hydrogen peroxidase will oxidise guaiac to form a blue-coloured compound (Benton *et al.*, 2015). A typical test kit consists of two small panels for a faecal smear and testing is usually performed using three cards (sampling three consecutive bowel movements) (Kuipers *et al*., 2013). When present in faeces, haem catalyses a chemical reaction which results in a blue colour appearing on the test paper. The threshold for positivity varies in different research or clinical settings. Some screening programmes refer subjects for colonoscopy when at least one panel test is positive, whereas others use thresholds of 5 to 6 panels (Kuipers *et al*., 2013).

Guaiac FOBT is one of the most studied and validated methods for CRC screening. Evidence for its usefulness comes from four large prospective, randomised trials published in the 1990s (Kronborg *et al*., 1996, Hardcastle *et al*., 1996, Kewenter *et al*., 1994, Mandel *et al*., 1993). These studies conducted in the UK, Sweden, Denmark and the USA offered gFOBT on an annual or biennial basis to individuals of different ages (ranging from 45 to 80 years old) and compared the results with non-screened controls. A systematic review combining those four trials (320,000 participants, 8 to 18 years of follow-up) showed a reduction in CRC-specific mortality by approximately 16% (Hewitson *et al*., 2008). Recent updates of the original trials have confirmed the long-term benefits of gFOBT screening in reducing CRC mortality (Shaukat *et al*., 2013, Scholefield *et al*., 2012).

Despite the fact that gFOBT is an affordable screening tool which has been proven to be beneficial, a major shortcoming of this method was evident in all trials: it had limited sensitivity for detecting cancer and poor performance in detecting adenomas (Kuipers *et al*., 2013). A study performing colonoscopies in individuals who had previous negative gFOBTs found CRCs in 4.5% and 8.6% of women and men, respectively (Schoenfeld

et al., 2005). Another study showed that gFOBT only detected 74% of colonoscopy-proven invasive cancers and 36% of advanced adenomas (Oort *et al*., 2010). Additionally, a Scottish trial demonstrated that *interval cancers*, those occurring after a negative result and before the next test date was due, comprised 31.2% of CRC cases after the first round of screening, 47.7% after the second and 58.9% after the third (Steele *et al*., 2012). This suggests that the sensitivity of gFOBT decreases as the screening is repeated and that some cancers may never bleed sufficiently to yield a positive result (Kuipers *et al*., 2013). Specificity is another concern. Firstly, haem is not specific to humans and there is a theoretical risk of false-positives caused by dietary factors. Secondly, haem is stable throughout the gastrointestinal tract, so a positive result might be due to bleeding in the upper portions of the gastrointestinal tract (Kuipers *et al*., 2013). As a result, the positive predictive value (PPV) for gFOBT varies from 0.9% to 18.7% for CRC and from 6% to 54.5% for adenomas (Hewitson *et al*., 2008). This means that a large proportion of patients will undergo colonoscopies unnecessarily. Furthermore, population compliance with gFOBT has been reported to be low, with uptake ranging from 34% to 55% in the community setting (Leuraud *et al*., 2013, Rees and Bevan, 2013). Despite these disadvantages, the strong evidence supporting its benefits and the low cost of the method has made gFOBT the preferred screening technique in many regions and countries (Benson *et al*., 2008).

Faecal immunochemical test (FIT)

FITs use antigen-antibody reaction to detect the presence of human globin in faeces. The test is more sensitive than the gFOBT since it detects smaller amounts of blood and allows for quantification of globin concentration in the faeces (Duffy *et al*., 2011, van Rossum *et al*., 2008). The FIT is theoretically also more specific for colonic bleeding because it detects only human blood and globin is not stable throughout the gastrointestinal tract. This means that bleeding from the upper portions of the digestive system is not likely to result in a positive test (Duffy *et al*., 2011). Several studies have

consistently proven that FIT performs better that its guaiac-based counterpart not only in terms of detection rate but also in population uptake (van Rossum *et al*., 2008, Hol *et al*., 2010, Oort *et al*., 2010).

However, FIT is more expensive than gFOBT and the samples are less stable at ambient temperature, thus requiring the test to be performed as promptly as possible (Duffy *et al*., 2011). In addition, the instability of globin in the gastrointestinal tract makes FIT less sensitive for detecting proximal colonic lesions (Haug *et al*., 2011). Most importantly, contrary to gFOBT, FIT-based strategies have not been evaluated in large prospective trials with adequate follow-up to provide survival data and their real capacity to reduce CRC deaths remains unclear. Notwithstanding these drawbacks, FIT is increasingly becoming preferred over gFOBT and this method has been accepted in many screening programmes worldwide (Allison *et al*., 2014).

Flexible sigmoidoscopy (FS)

Guaiac-FOBT was the only screening method supported by prospective randomised trials until the late 2000s, when several studies conducted in Europe and in the USA demonstrated the benefit of FS, the use of a flexible endoscope to image de distal portion of the colon, as a screening tool (Segnan *et al*., 2011, Atkin *et al*., 2010, Schoen *et al*., 2012, Hoff *et al*., 2009, Thiis-Evensen *et al*., 1999). The procedure is performed without sedation and with simple bowel preparation using an enema. A meta-analysis/systematic review evaluated the results from all those trials, and concluded that FS reduces the incidence of CRC by 18% and mortality by 28% in an intention-to-treat analysis. When analysing only those subjects who actually adhered to the screening programme, reductions of 32% and 50% in incidence and mortality respectively were observed (Elmunzer *et al*., 2012).

The use of FS as a population-based screening strategy has some major problems. Firstly, it requires proper facilities, trained personnel and has higher costs than gFOBT and FIT (Kuipers *et al*., 2013). Secondly, it is

associated with a lower population uptake than the previously cited methods (Hol *et al*., 2010, Atkin *et al*., 2010, Segnan *et al*., 2011). Lastly, it does not allow evaluation of the proximal colon, thus potentially missing some lesions (Schoenfeld *et al*., 2005). For this reason, many experts consider colonoscopy a better option for endoscopic screening (Kuipers *et al*., 2013).

Colonoscopy

Colonoscopy is a diagnostic procedure which uses a flexible video-endoscope to image the whole colon. It is usually performed with complete bowel cleansing and varied degrees of sedation. Colonoscopy is widely used as a primary screening method in several countries such as the USA, Canada, Germany and Poland (Kuipers *et al*., 2013). The first study to provide strong evidence supporting the use of colonoscopy for CRC screening was the case-control National Polyp Study in the USA. This trial included patients who had undergone colonoscopic polypectomy and were followed up carefully. The authors reported a reduction in the incidence of CRC by 76-90% (Winawer *et al*., 1993). Later, a long-term follow-up report suggested a 53% reduction in CRC mortality (Zauber *et al*., 2012). A reduction in CRC incidence ranging from 11% to 19% (depending on sex- and age-groups) was also demonstrated in studies analysing the results of the German nationwide colonoscopy-based screening programme (Brenner *et al*., 2009, Pox *et al*., 2012).

The main drawbacks for the widespread use of colonoscopy are its costs, availability, invasiveness and complications. Colonoscopy is one of the most expensive screening tests available and demands well-equipped facilities and highly trained teams, thus limiting its use for deprived communities (Gupta *et al*., 2014). Even when available, the invasive aspect of the test, associated with the need for bowel preparation results in cumulative population uptake rates as low as 15-17% as demonstrated in the German nationwide programme (Pox *et al*., 2012). Complication rates in colonoscopies have declined recently and population based studies have shown rates of serious complications or perforations of 0.058% to 0.082%

(Pox *et al*., 2012, Arora *et al*., 2009). However, complication rates vary considerably in small practice facilities in community settings and can be as high as 0.63% (Arora *et al*., 2009). Another negative aspect of colonoscopy is the fact that this procedure also has a lower sensitivity for small lesions and those located in the proximal colon in the same way that happens with FS (Kuipers, 2014). Although widely accepted as a screening method, colonoscopy is not an ideal tool and, therefore, new strategies are being assessed to improve the accuracy and cost-effectiveness of CRC screening.

CT-colonography

Although not as thoroughly tested as the previous methods, CT-colonography is an accepted method for CRC screening in some regions, particularly the USA (Levin *et al*., 2008). CT-colonography (or virtual colonoscopy) uses computed tomography technology to image the entire colon with the aid of iodinated contrast agents and bowel distension. It requires less bowel preparation and can detect 70-100% of the advanced neoplasms (advanced polyps and invasive cancers) that are visualised by colonoscopy (Johnson *et al*., 2008, Kim *et al*., 2007a, Graser *et al*., 2009, Stoop *et al*., 2012). It had higher population uptake when compared with colonoscopy in a study setting (Stoop *et al*., 2012). Nonetheless, contrary to the initial expectations, patients complained more about the burden of CT-colonography than colonoscopy, particularly the disturbed bowel movements that were caused by the iodine contrast agent. In addition, CT-colonography is less cost-effective than other screening modalities, thus limiting its application as a widespread screening method (Kuipers *et al*., 2013).

1.5.2. **Colorectal cancer screening practices in the UK and in Brazil**

In the early 2000s, a pilot study was conducted in the UK to analyse the feasibility of introducing a population-based screening programme for CRC. It was performed in regions of Scotland and England and was based on the UK gFOBT study (Hardcastle *et al*., 1996). More than 478,000

residents from the designated areas aged 50-69 years were invited to participate and the uptake was 56.8%. The overall positive rate was 1.9% and the rate of cancer detection was 1.62 per 1000 people screened (UK Colorectal Cancer Screening Pilot Group, 2004). The complication rate in those undergoing colonoscopy was low (0.56%) as was the perforation rate (0.05%), with no deaths being attributable to the procedure. Each site subsequently reported results from the second and third rounds of the screening test and this confirmed the feasibility and safety of the strategy. The programme costs £5,900 per life-year saved, which was considered to be acceptable for a screening programme (Rees and Bevan, 2013). Based on these results, in 2006 a national biennial gFOBT screening programme was initiated in the UK, including individuals 60-69 years-old. The age limit was further extended to 60-74 years in 2010. Since then, almost 15,000 CRC cases have been identified by screening with a clear predominance of early-stage tumours compared to non-screen-diagnosed cancers. This has resulted in better survivals as depicted in figure 1.16 (Rees and Bevan, 2013). More recently, FS was also added as a screening option for individuals at age 55 in the UK as a result of several studies showing the benefit of this strategy in reducing incidence and/or mortality due to CRC (Segnan *et al*., 2011, Atkin *et al*., 2010, Schoen *et al*., 2012, Hoff *et al*., 2009).

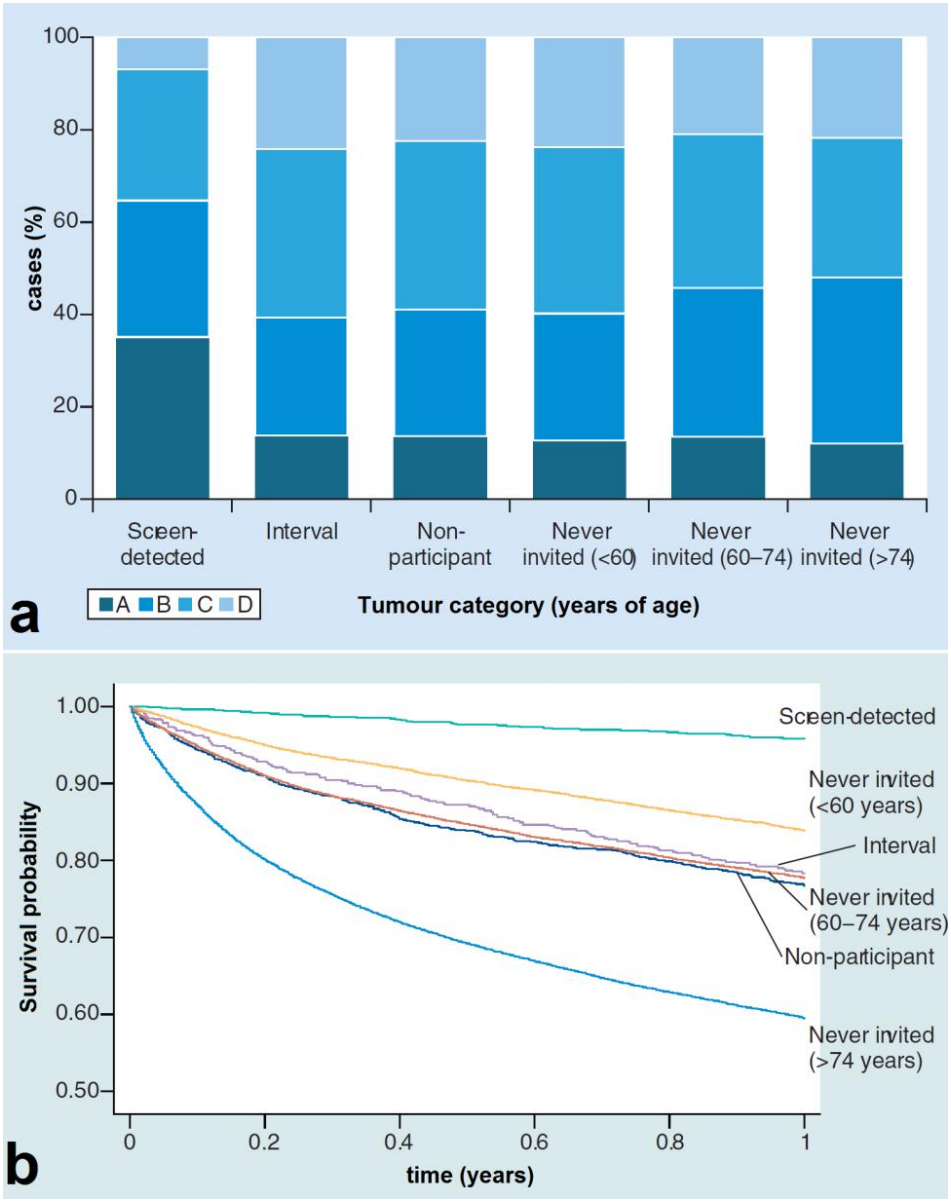


Figure 1.17. Comparison between screen-detected versus non-screen-detected cancers. a) a shift from late stages (Dukes' stages C and D) to predominantly early stage tumours (Dukes' stages A and B) is observed when the diagnosis is made during a screening test; b) the corresponding effect of early diagnosis on survival. Adapted from (Rees and Bevan, 2013).

In Brazil, the Unified Health System ("*Sistema Unico de Saúde"* - SUS) provides health care for the entire Brazilian population. However, approximately 20% of the population uses private health insurance for most of their medical needs resulting in very divergent standards of care (Pinto and Soranz, 2004). Even so, there is no large scale CRC screening programme in Brazil in either the public or the private sectors. Literature

addressing CRC screening in Brazil is scarce and most reports come from small studies (Perez *et al*., 2008, Altenburg *et al*., 2007). One of the largest trials invited part of the population over 40 years-old from a small city to undertake gFOBT. In total, 4,567 tests were delivered and the uptake was approximately 80%. The positivity rate was 10.7% and these patients were further invited to undergo colonoscopy. Fifty nine polyps and 9 carcinomas were identified, resulting in incidences, in the screened population, of 1.3% and 0.2%, respectively (Perez *et al*., 2008).

CRC screening is advised by Brazilian medical societies (Assis, 2011) and even by the government-sponsored National Cancer Institute. However, as cited earlier, no official publicly- or privately-funded programme is currently being conducted within the country. As a result, particularly for the 80% of population relying exclusively on SUS, most CRC cases are diagnosed at late stages (Valadão *et al*., 2010) with a corresponding increase in the mortality rate and the economic burden for the nation (Torres *et al*., 2010).

1.5.3. **New strategies under research**

Based on the evidence described, it is clear that the currently used tools are far from fulfilling the requirements for a good screening method. Consequently, there are several new approaches under investigation aiming to develop more successful strategies for CRC screening (Miller and Steele, 2012, Di Lena *et al*., 2013, Bosch *et al*., 2011, Nguyen and Weinberg, 2016). Most of these methods are based on the identification of cancer markers in faeces or blood. The presence of cancer markers in the faeces occurs due to leakage of substances into the intestinal lumen (e.g. haemoglobin), active secretion by luminal cells (mucins, etc.) or exfoliation of cells to the lumen carrying cellular components such as proteins, DNA and RNA. The mechanisms responsible for the translocation of cancer markers into the bloodstream are less clear and may involve circulating tumour cells,

macrophage release or tumour-derived plasma microvesicles (Bosch *et al*., 2011).

Faecal markers

The pursuit of CRC markers in faeces has some theoretical advantages. Tumours exfoliate more than the normal epithelium, resulting in large numbers of malignant cells in the sample. Also, DNA from normal colonocytes is degraded during apoptosis (which precedes cell shedding), whereas tumour cell-derived DNA suffers less degradation due to the anti-apoptotic properties of cancer cells, thus resulting in longer and more stable DNA (Bosch *et al*., 2011).

The current methods for detecting DNA alterations in faeces screen for mutations, aberrant methylation or both. Since some genes are frequently mutated in CRC, several analyses have been undertaken based on the identification of DNA mutations in faeces. Mutations in *KRAS, TP53, APC* and others have been used individually or in combination, resulting in low accuracies (Miller and Steele, 2012, Di Lena *et al*., 2013, Ahlquist *et al*., 2000). Attempts to improve these results were made by associating mutation patterns with MSI markers and the DNA integrity assay (DIA), a marker for long DNA typical of exfoliated tumour cells. The first assay to combine *KRAS*, *TP53* and *APC* mutations with a MSI marker (*BAT26*) and DIA showed high sensitivity (91% for carcinomas and 82% for adenomas larger than 1cm) and high specificity (93% for both) (Ahlquist *et al*., 2000). Unfortunately, later studies have not reproduced these results and lower sensitivities have been reported. In a comparative study using a panel of 21 genes, Imperiale *et al*. showed a sensitivity of 52% for the DNA test compared with 13% for FOBT for detecting invasive cancers, whereas both tests exhibited poor sensitivities for detecting advanced adenomas (18% and 11% for the DNA test and FOBT respectively) (Imperiale *et al*., 2004). The use of methylation patterns has also been explored since the discovery of the hypermethylator phenotype in CRCs. Screening for methylation in specific CpG islands resulted in a sensitivity of 38-94% for CRC (21-70% for adenomas) and a specificity of 79-

100% (Bosch *et al*., 2011). The combination of both strategies seems to be the most promising. Recently, Imperiale *et al*. reported the results of a study comparing FIT with a multitargeted panel combining the following markers: *KRAS* mutation, aberrant *NDRG4* and *BMP3* methylation, *ACTB* plus a haemoglobin immunoassay. Sensitivities for advanced precancerous lesions were 47.4% for the stool DNA test and 23.8% for FIT. For invasive cancers, sensitivities were 92.3% and 73.8%, respectively. However, specificities were 86.6% and 94.9%, favouring FIT (Imperiale *et al*., 2014). Despite the lack of long term studies showing benefit in reducing mortality, several faecal DNA tests are commercially available and this technique is currently accepted for screening purposes in the USA (Levin *et al*., 2008).

Faecal RNA has also been used as a screening tool. *MMP7* (encoding matrix metalloproteinase-7) and *PTGS2* (encoding prostaglandin G/H synthase 2) have been assessed in small case-control studies and have shown conflicting results (Bosch *et al*., 2011). The use of a larger panel (9 genes) has shown more promising results with a sensitivity of 78% and a specificity of 100%. However, the sample sizes in this and other similar studies have been very small and these results must be validated in larger cohorts (Bosch *et al*., 2011). MicroRNAs – short, non-coding 18-22 nucleotide RNAs, have also been evaluated for screening purposes (Miller and Steele, 2012, Di Lena *et al*., 2013). Several miRNAs are over-expressed in cancer and this seems to be a promising technique for early diagnosis, although it is still at a very early stage of development (Di Lena *et al*., 2013).

Apart from haemoglobin, various proteins (such as calprotectin, CEA, lysozyme, albumin, α1-antitrypsin, MUC1, lactoferrin, α-defensin 1, DAF, M2-PK, and CLU) have been assessed in faeces from CRC patients. None of them is however sensitive and specific enough to be adopted for screening purposes and many have cross-specificity with inflammatory bowel disease (Bosch *et al*., 2011).

Blood markers

The use of faecal tests has faced low acceptance by the population as demonstrated by the screening trials and population-based programmes discussed earlier. Venepuncture, although somewhat invasive, is less inconvenient than manipulating, storing and mailing stool (Bosch *et al*., 2011). Furthermore, the majority of the target population for CRC screening is likely to have blood tests for other purposes periodically, making it even easier to incorporate a new screening test.

Studies analysing the presence of DNA mutations in blood shortly followed the first reports of DNA mutations in the stool of CRC patients. In general, the same target mutations (mainly *KRAS* and *APC*) have been analysed in the blood as have been in stool samples. Sensitivities for DNA mutations in blood were lower when compared with the faecal analysis, and no improvement was observed when combining mutations and aberrant methylation (Bosch *et al*., 2011). The use of isolated methylation patterns has also been tested. A promising marker, the aberrant methylation of *SEPT9*, was recently tested against FIT in a colonoscopy-controlled trial. The authors reported a sensitivity of 73% for the *SEPT9* test and 68% for FIT, while the specificities were 81.5% and 97.4%, respectively (Johnson *et al*., 2014). The lower specificity for DNA-based tests both in faecal and blood samples raises concerns about the real meaning of false-positive results. They could represent an inherent lack of specificity of the test or occult neoplasms that are not yet detectable by colonoscopy.

Blood RNA expression analysis is another alternative method under investigation. Most studies have focused on *CEA*, *CK19* and *CK20* expression for CRC detection. Although variations in study design prevent direct comparisons between results, sensitivities have been reported between 8%-75% for individual markers and 60%-89% for combinations (Bosch *et al*., 2011). Specificities have ranged from 78% to 100%. Micro RNAs are more stable in plasma than mRNAs due to their small size and associated proteins and many have been tested in blood samples as potential cancer screening markers (Slaby *et al*., 2009, Yi *et al*., 2016). For

example, miR-92 expression was able to discriminate between CRC patients and normal controls with a sensitivity and specificity of 89% and 70%, respectively (Ng *et al*., 2009). Combined use of miR-29a and miR-92a showed improved specificity, while maintaining the same sensitivity (Huang *et al*., 2010). The use of genome-wide expression profiles is also being evaluated. Although some promising results have been produced, more research is still needed to clarify the importance of this technology for CRC screening (Bosch *et al*., 2011).

The emerging field of proteomics has increased the interest in the study of proteins as biomarkers of diseases. New technologies such as matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) and surface-enhanced laser desorption/ionization (SELDI-TOF) mass spectrometry, permit the identification of peptide patterns that can differentiate serum from CRC patients and control individuals (Bosch *et al*., 2011). However, no protein marker has shown an adequate performance to be used for CRC screening until now. The most studied and used protein marker for the evaluation of CRC patients is carcinoembryonic antigen (CEA). Although CEA is useful for post treatment surveillance, the accuracy of CEA for screening is dismal (Tan *et al*., 2009). Other proteins have however exhibited promising results such as: sCD26 (90% sensitive, 90% specific), Alpha-defensin 1 (69% sensitive, 100% specific), Laminin (89% sensitive, 89% specific), CCSA-2 (89%-97% sensitive, 78%-84% specific), TIMP-1 (60% sensitive, 98% specific), and Clusterin (56% sensitive, 100% specific) (Bosch *et al*., 2011). None of them, however, has been validated to date for screening use. Using a proteomic approach (SDS-PAGE and MALDI-TOF), researchers selected CRMP-2 among 325 proteins in secretomes of cell lines from 12 cancer types. Testing the protein level in plasma from 201 CRC patients and 201 healthy controls (associated with the level of CEA), they found a sensitivity of 77% and a specificity of 95% for this protein (Wu *et al*., 2008). Therefore, the detection of blood-borne proteins for CRC screening holds great promise, although it needs to be further studied and new markers are highly necessary.

1.6. Proteomic dissection of Wnt activation models and the discovery of novel candidate biomarkers

A biomarker is defined as “a characteristic used to measure and evaluate objectively normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention” (Atkinson *et al*., 2001). In practice, biomarkers are used for diagnosing diseases, predicting clinical outcome (prognostic biomarkers), predicting response to specific treatments (predictive biomarkers) or for surveillance after treatment (Pritzker, 2015, Atkinson *et al*., 2001).

No protein has yet been validated as a CRC biomarker for screening purposes, and very few markers are currently used for prognostic and predictive reasons (Gonzalez-Pons and Cruz-Correa, 2015). Carcinoembryonic antigen (CEA) is the only accepted marker for detecting recurrence after treatment of CRC, although its performance remains suboptimal (Tan *et al*., 2009). However, the study of colorectal cancer proteomics both in human and animal models has provided important pieces of evidence about CRC carcinogenesis. Several candidate biomarkers have been suggested using this approach (de Wit *et al*., 2013, Alvarez-Chaver *et al*., 2014). For example, Leclerc *et al*. recently reported an analysis of an intestinal tumourigenesis model induced by low dietary folate in methylenetetrahydrofolate-reductase deficient mice (*Mthfr*$^{+/-}$) (Leclerc *et al*., 2014). They showed a higher expression of several members of the Nurd (nucleosome remodelling and histone deacetylation) complex, Kras, Sumo (small Ub-like modifiers)-related proteins, heat shock proteins and fatty acid metabolism-related proteins. In another study, de Wit *et al*. reported the differential proteomic profiling of the CRC secretome, using CRC samples compared to adjacent normal mucosa, and CRC cell lines. They found 409 proteins that were more abundant in CRC versus normal tissues (de Wit *et al*., 2014). Based on consistency and abundance, and after performing a clustering analysis, they selected 7 potential candidate biomarkers to be further assessed in future studies: SSRP1, SUP16H, RRM1, DNMT1, MCM3, MCM5, MCM6 and TRIM28.

Nevertheless, few of the candidate biomarkers that have been reported in proteomic studies have been assessed in human blood (Alvarez-Chaver *et al*., 2014). In addition, most recent studies in CRC have not focused on specific genetic pathways. Conversely, they have screened for general differential expression patterns between cancer tissues and adjacent mucosa. As most of the genetic alterations present in CRC (and the resulting possible alterations in protein expression) are “passengers” instead of “drivers” of the carcinogenic process (Wood *et al*., 2007), a more focused proteomic analysis targeting specific pathways known to be involved in CRC development would be expected to yield more promising biomarker candidates.

As the Wnt pathway is the most important driver of the early stages of colorectal tumourigenesis, the proteomic profiling of Wnt activation-based models is likely to provide candidate biomarkers which could ultimately translate into useful clinical tools for CRC screening, prognosis or prediction of response to treatment. A study with this purpose has been previously conducted in our group by Hammoudi *et al*. who performed a proteomic dissection using two different murine models of *Apc* inactivation (Hammoudi *et al*., 2013). One of them was the *AhCre*$^{+}$ *Apc*$^{fl/fl}$ mouse, an animal bearing loxp-flanked *Apc* alleles and a *Cre*-recombinase transgene. When this animal is exposed to β-naphthoflavone injections, Cre-recombinase transcription is activated and this results in the deletion of the loxp-flanked *Apc* alleles specifically in the intestinal epithelium, thus causing an acute activation of the Wnt pathway (Sansom *et al*., 2004). The other model used was the *Apc*$^{Min/+}$ mouse. This animal has a germline mutation in one *Apc* allele simulating a FAP patient, and spontaneously develops multiple intestinal neoplasms (Min) during its life-span (Su *et al*., 1992, Moser *et al*., 1990). Both models were compared with *Apc*$^{+/+}$ animals of the same genetic background. Analysing epithelial cell extracts and intestinal tissues, the authors found 81 proteins to be up-regulated in *Apc*$^{fl/fl}$ versus *Apc*$^{+/+}$ mice. Ingenuity Pathway Analysis (IPA) suggested that 13 of these proteins could also be detected in blood or serum. Four were excluded due to low specificity for CRC or reagent issues. The 9 remaining candidates were further assessed using

immunohistochemistry, Western Blotting and qPCR. After these steps, 6 promising candidates were selected for ELISA (using mouse serum) and for testing in human CRC tissues and adjacent normal mucosa using qPCR. Finally, four candidate proteins were suggested as potential biomarkers in CRC: Nucleolin (NCL), Prohibitin (PHB), Cytokeratin 18 (CK-18) and Ribosomal protein L6 (RPL6) (Hammoudi *et al*., 2013). Further unpublished work performed in our group has reinforced the high expression of these biomarkers (Dr Fei Song, currently working at Infrafrontier GmbH, Munich, Germany). Additionally, it has expanded the candidate list by adding Nucleosome assembly protein 1 – like 1 (NAP1L1), Splicing factor arginine/serine-rich 2 (SFRS2), Fatty acid binding protein 6 (FABP6), Nucleophosmin, and DEAD box protein 5 (DDX5) to the group of candidate proteins (John R Jenkins, Personal Communication).

To further refine the selection of the candidate biomarkers, a colleague, Dr Shahram Ali Ibrahim, carried out additional studies to validate the expression patterns of these candidate proteins using animal models, cell lines and human CRC samples from the Countess of Chester Hospital NHS Foundation Trust (Ibrahim, 2014). His work identified NAP1L1, RPL6, HMGB1 and Prohibitin as the most promising candidates. Furthermore, based on a report that showed that CDC5L over-expression displaces SFRS2 from nuclear speckles (Engemann *et al*., 2002), Ibrahim´s work also suggested antagonistic actions between these proteins during Wnt activation, which could trigger an apoptotic burst. As a result of all the work mentioned above, a final list of candidate biomarkers was selected for further studies focusing on human tissue and blood samples. These proteins will be evaluated throughout the present project and are therefore briefly described in the next section.

1.6.1. **Novel candidate biomarkers**

Nucleosome assembly protein 1 – like 1 (NAP1L1)

NAP1 is a protein involved in nucleosome assembly and disassembly and was initially discovered in HeLa cells (Park and Luger, 2006). The nucleosome is the basic unit of chromatin and is formed in a 2-step process: a tetramer of the histone subunits H3 and H4 is first deposited onto the DNA. Then, the process is completed by the addition of 2 heterodimers of the H2A and H2B subunits. NAP1L1 is the human counterpart of yeast NAP1, and is thought to work as a histone "chaperone", carrying histone subunits during chromatin remodelling (Park and Luger, 2006). Other suggested functions of NAP1L1 include roles in gene transcription and cell cycle regulation (Park and Luger, 2006). The protein localises predominantly in the cytoplasm (in HeLa cells), with a low but definitive amount also being detected in the nucleus (Okuwaki *et al*., 2010). Studies using murine pluripotent stem cells have suggested that NAP1L1 is important in maintaining the undifferentiated state, and *NAP1L1* gene knockdown in these cells causes differentiation into cardiomyocytes and mesoderm development (Gong *et al*., 2014, Li *et al*., 2012).

There is also some evidence that NAP1L1 is involved in carcinogenesis. One study showed that *NAP1L1* expression was increased in small intestinal neuroendocrine (carcinoid) tumours compared with colorectal adenocarcinomas and normal adjacent tissues (Kidd *et al*., 2006). Another author demonstrated that silencing the *NAP1L1* gene resulted in decreased tumour proliferation in pancreatic neuroendocrine cancer cell lines and xenografts (Schimmack *et al*., 2014). Additionally, *NAP1L1* was shown to be over-expressed in CRC and hepatoblastomas (Line *et al*., 2002, Nagata *et al*., 2003). However, the literature assessing whether NAP1L1 is a potential cancer biomarker remains poor.

Ribosomal protein L6 (RPL6)

Ribosomes, the fundamental organelles responsible for protein biosynthesis, are made up of two components: ribosomal RNA (rRNA) and ribosomal proteins (RPs). There are around 80 different RPs, whose function involves the stabilisation of specific rRNA structures in ribosomal subunits and the promotion of correct folding of rRNAs during ribosomal assembly (Lai and Xu, 2007). In addition, RPs exhibit various poorly understood extra-ribosomal functions. Several different RPs have been shown to be over-expressed in some types of cancer such as breast, prostate, oesophageal, cervical and hepatocellular tumours (Wu *et al*., 2011). However, RPL6, the candidate biomarker suggested from the work described previously, has only been associated with one type of cancer (gastric neoplasms) and has been addressed to a limited extent in the scientific literature (Du *et al*., 2005, Gou *et al*., 2010, Wu *et al*., 2011). Studies have shown that RPL6 expression is associated with cell proliferation and growth in gastric cancer cell lines. Additionally, it has been suggested that the mechanism by which RPL6 could affect these processes is via Cyclin E regulation in the cell cycle (Wu *et al*., 2011, Gou *et al*., 2010) and regulation of the Bcl-2/Bax equilibrium for apoptosis (Du *et al*., 2005). In addition, using multidrug resistant cell line models, the authors demonstrated that RPL6 is important for the prevention of apoptosis caused by cytotoxic drugs and that *RPL6* knockdown can sensitise cells to the destructive effects of those compounds (Du *et al*., 2005). It has been suggested that RPL6 over-expression in gastric cancer samples could be a negative prognostic factor resulting in reduced survival (Wu *et al*., 2011). To the best of our knowledge, apart from Hammoudi´s work previously discussed, no study has reported the role of RPL6 in colorectal tumours.

Prohibitin (PHB)

The *PHB* gene is located at the chromosomal position 17q21 and encodes a protein that is ubiquitously expressed and present in multiple

cellular compartments (Zhou and Qin, 2013). It is associated with several biological functions depending on its localisation. In the nucleus, PHB regulates the cell cycle, senescence mechanisms and causes tumour suppression (Zhou and Qin, 2013). In mitochondria, it is involved in the control of oxidative stress (Thuaud *et al*., 2013, Chowdhury *et al*., 2017), protein synthesis (He *et al*., 2012), maintenance of mitochondrial DNA copy number (Kasashima *et al*., 2008), mitophagy (Galluzzi *et al*., 2017) and transport of newly synthesised proteins (Nijtmans *et al*., 2000). In the plasma membrane, it acts as a lipid scaffold to maintain membrane integrity (Thuaud *et al*., 2013, Zhou and Qin, 2013). It has been suggested that PHB regulates cell proliferation and growth through the interaction with tumour suppressors such as p53 and RB, and proto-oncogenes such as MYC (Zhou and Qin, 2013). Furthermore, recent studies have established that PHB activates the Ras-Raf-MEK-Erk signalling cascade, an important pathway associated with cell proliferation and survival (Chowdhury *et al*., 2014). This results in enhancement of Bcl-2 expression and decreases in both Bax-Bak activation and cytochrome c release from mitochondria, causing inhibition of apoptosis. Supporting this concept, a study by Kim *et al*. showed that *PHB* knockdown using siRNA enhanced sensitivity to anthralin-mediated cell death in human keratinocyte cell lines (Kim *et al*., 2007b).

In cancers, there have been contradictory findings regarding PHB expression and significance. In ovarian cancer, a study showed reduced PHB expression in cancer tissues compared with normal specimens (Jia *et al*., 2014). Reduced expression of PHB has also been found in hepatocellular and cholangiocarcinoma cells (Fan *et al*., 2016). In contrast, PHB over-expression has been demonstrated in lung (Jiang *et al*., 2013, Guo *et al*., 2012), breast (Najm *et al*., 2013), gastric (Kang *et al*., 2008), thyroid (Franzoni *et al*., 2009), prostate (Ummanni *et al*., 2008), superficial (Wu *et al*., 2007) and invasive bladder cancers (Cao *et al*., 2016). Among these studies, the few which reported clinical correlation showed that increased expression of PHB was associated with a better prognosis for breast and superficial bladder cancer, whilst it was deleterious in lung (poor survival), gastric (poor differentiation) and invasive bladder (poor survival) tumours.

Noteworthy, Kang *et al*. also tested serum from gastric cancer patients and healthy controls using ELISA and did not find any difference in the concentration of PHB between these groups (Kang *et al*., 2008). In colorectal cancer patients, Chen *et al*. demonstrated an increased expression of PHB in cancer tissues compared to normal matched mucosa and adenomas (Chen *et al*., 2010a). No difference was however observed between normal samples and adenomas. In this study, although a positive correlation was found between PHB over-expression and poor differentiation, no association was demonstrated with tumour stage or survival.

High mobility group box 1 (HMGB1)

HMGB1 is a remarkably multifunctional protein which possesses diverse biological functions (Ohmori *et al*., 2011). A comprehensive review about its biology, also covering other members of the family, has recently been published (Kang *et al*., 2014). This group of proteins (HMGs) was named due to the high mobility pattern showed in polyacrylamide gels. Interestingly, the same ability is seen in the intracellular movement of the protein. HMGB1 was first isolated from calf thymus and is predominantly located in the nucleus of cells. Nonetheless, the protein can leave the nuclear space in less than 2 seconds and is able to shuttle to the cytoplasm and even to the extracellular space (Kang *et al*., 2014). Inside the cell, HMGB1 is involved in DNA replication, repair and recombination, as well as transcription and genomic stability. The protein can translocate to the extracellular matrix either through passive release (from dead or injured cells) or active secretion (from immune and cancer cells). Thereafter, HMGB1 can bind to several cell receptors such as the receptor for advanced glycation end products (RAGE) and Toll-like receptors (TLR). When bound, the complex induces chemotaxis, migration, proliferation and differentiation of tumour and immune cells (Kang *et al*., 2013). These effects have an important role in inflammatory and immune diseases.

As a result of its complex functions, HMGB1 has both pro and anti-tumour effects in cells expressing RAGE (Ohmori *et al*., 2011, Kang *et al*.,

2013). Some of the pro-tumourigenic properties of HMGB1 are: a) sustenance of an inflammatory environment; b) changes in energy metabolism; c) promotion of invasion and metastasis; d) inhibition of anti-tumour immunity, and e) promoting angiogenesis. By contrast, HMGB1 can also prevent tumours via: i) interaction with tumour suppressors; ii) increase of genome stability, and iii) increase of autophagy (Kang *et al*., 2013). In the last 20 years, HGMB1 has been quite extensively studied in cancer. It has been associated with carcinogenesis in colon, breast, lung, prostate, cervical, skin, kidney, stomach, pancreatic, liver, bone and blood cancers (Kang *et al*., 2014). In CRC, a recent study assessing tumour tissues using immunohistochemistry showed that 56% of samples expressed high levels of HMGB1 and this over-expression was associated with adverse prognostic factors such as tumour stage and differentiation (Süren *et al*., 2014). Other research demonstrated that tumours with both nuclear and cytoplasmic expression of HMGB1 had less lymphocyte infiltration and a decreased 5-year survival rate (Peng *et al*., 2010). Lee *et al*. evaluated HMGB1 levels in sera from patients with CRC and healthy controls and compared the findings with another serum biomarker, carcinoembryonic antigen (CEA). For diagnostic purposes, HMGB1 results exhibited a better specificity compared to CEA (96% vs 90.7%); however, the sensitivity was worse (20.1% vs 25.6%). The authors suggested that both tests could be used in combination, since HMGB1 performed slightly better in early stage lesions whereas CEA had better accuracy for late stage tumours (Lee *et al*., 2012).

Splicing factor arginine/serine-rich 2 (SFRS2) or Serine/arginine-rich splicing factor 2 (SRSF2) or SC35

Serine/arginine rich (SR) proteins are a family of RNA binding proteins characterised by RNA recognition motifs and a signature SR domain enriched with serine and arginine repeats (Fu, 1995). SR proteins play an essential role in both constitutive and alternative splicing (Fu, 1995, Graveley, 2000, Manley and Tacke, 1996, Anko, 2014). SFRS2 (also called SRFS2 or SC35 – it will be referred solely to as SFRS2 hereafter), has been

particularly associated with genomic stability, cell proliferation and organogenesis in mammalians (Xiao *et al*., 2007). Studies have shown that SFRS2 is also involved in transcriptional elongation (Lin *et al*., 2008), mRNA stabilisation (Qian *et al*., 2011) and regulation of apoptosis (Merdzhanova *et al*., 2008). The importance of alternative splicing for cell physiology is highlighted by the fact that more than 90% of human genes undergo this alternative processing, resulting in diverse protein isoforms (Wang *et al*., 2008, Pan *et al*., 2008). Mounting evidence suggests that alternative splicing is also an important driving force for carcinogenesis, interfering with oncogene function, signalling cascades and response to the tumour microenvironment (Biamonti *et al*., 2014).

Few studies have addressed the role of SFRS2 in cancer models. Studying lung cancer cell lines and tumour xenografts, Merdzhanova *et al*. showed that SFRS2 is important in the control of pro-angiogenic and anti-angiogenic isoforms of VEGF-A (Merdzhanova *et al*., 2010). Edmond *et al*. demonstrated that post-translational modifications of SFRS2 are involved in the control of cell fate in response to the cytotoxic agent cisplatin (Edmond *et al*., 2011). In head and neck cancer cell lines, Sharma *et al*. have shown that SFRS2 causes E-cadherin mis-splicing and transcript degradation, thus decreasing the expression of this protein - a feature associated with metastatic transformation (Sharma *et al*., 2011). We have not been able to find any published study evaluating the role of SFRS2 in colorectal cancer.

Cell division cycle 5-like (CDC5L)

CDC5L, the human homolog of the product of the *cdc5*$^{+}$ gene in *Schizosaccharomyces pombe*, is an essential protein involved in the composition of a larger multi-protein complex which is part of the spliceosome (Ajuh *et al*., 2000). This structure is responsible for processing the pre-mRNA into mature mRNA. The importance of CDC5L for mRNA processing is supported by experiments showing that its depletion causes ineffective splicing, although the spliceosome itself is still formed (Ajuh *et al*., 2000). Apart from this well-established function, there is also evidence that

CDC5L is important for cell cycle progression (Bernstein and Coughlin, 1998). A study demonstrated that different patterns of CDC5L phosphorylation resulted in distinctive mRNA processing and that the protein might be a target for cyclin-dependent kinase 2 (CDK-2), a factor associated with S-phase progression (Gräub *et al*., 2008). Another possible mechanism for cell cycle control involves “checkpoint” actions. It has been shown that CDC5L interacts with the cell-cycle checkpoint kinase ataxia-telangiectasia and Rad3-related protein (ATR) and the resulting complex works by halting cell proliferation in response to DNA damage (Zhang *et al*., 2009). Additionally, recent research showed that CDC5L knockdown in HeLa cells caused dramatic mitotic arrest and chromosomal misalignment, eventually leading to what is called mitotic catastrophe (Mu *et al*., 2014).

There are few studies assessing the role of CDC5L in cancer. Although there is one study reporting its over-expression in cervical cancer (Mu *et al*., 2014), most of the literature refers exclusively to osteosarcoma, a bone cancer typically found in children and young adults (Mu *et al*., 2014, Lu *et al*., 2008, Martin *et al*., 2014). Initial reports in the late 2000s suggested that CDC5L might be the oncogene involved with the amplification of region 6p12-p21 frequently observed in osteosarcomas (Lu *et al*., 2008). Recently, another study analysed the expression of 17 candidate genes in osteosarcoma patients aiming to assess their potential for predicting response to preoperative (neoadjuvant) chemotherapy. Their results confirmed that CDC5L, RUNX, CDK4 and RECQL4 expressions are correlated with poor response to treatment (Martin *et al*., 2014). Apart from these studies, we were not able to find any other evaluation of CDC5L in other cancer types, including CRC.

1.7. Hypothesis

The candidate proteins selected from the proteomic dissection of animal models of colorectal carcinogenesis based on *Apc* inactivation may be useful colorectal cancer biomarkers for early diagnosis, prognosis or prediction of response to treatment. Furthermore, mechanistic studies assessing these candidates may provide important new data about the biology of CRC.

1.8. Aims of the study

a. To assess the expression of candidate proteins previously selected in our research group using human clinical samples from individuals without intestinal neoplastic lesions, with colorectal polyps and with colorectal cancer, thus encompassing the entire normal-adenoma-carcinoma sequence;

b. To explore the gene expression pattern of these candidates in normal and neoplastic colorectal tissues;

c. To test the concentrations of promising candidate proteins in the blood of individuals with and without CRC;

d. To identify possible mechanisms by which these biomarkers are involved in intestinal carcinogenesis;

e. To investigate whether there is any association between the expression of these candidate proteins and prognosis in CRC patients;

f. To compare, whenever possible, the results obtained in two populations with diverse ethnic, geographic and social backgrounds (UK and Brazil), thus increasing the validity of any conclusions.

Chapter Two:

Patients, materials and methods

2. CHAPTER 2 – PATIENTS, MATERIALS AND METHODS

2.1. Ethical approval

All samples from the prospective “Brazilian cohort” were collected after the nature of the research had been explained to individuals, its potential benefits and risks had been discussed, and “informed consent” had been confirmed. The research project was approved by the ethical boards listed below, including the retrospective collection of samples and data for the “prognostic cohort”. A research code was assigned to each subject, so that direct identification of the person was not possible throughout the work. This research was approved by the Committee for Research Ethics of the Hospital Universitario Julio Muller - Federal University of Mato Grosso, Cuiaba – Brazil, and by the Brazilian National Commission for Research Ethics (CONEP), decision number: 1.628.901.

Samples from the UK cohort were also collected with informed consent and following research ethics committee approval (NREC number 12/NW/0011).

2.2. Brazilian samples

a. Cancer patients

The “Brazilian cohort” of samples was prospectively collected in the city of Cuiaba, capital of the state of Mato-Grosso, Brazil. The author of this thesis participated directly in the process of sample collection along with a colleague (Dr Lenuce Ydy, Surgical Oncologist). Clinicians from different hospitals and cancer centres were asked to identify patients with a recent diagnosis of CRC for whom a surgical resection of the tumour was scheduled. Then, the patient was interviewed, the nature of the research was explained and they were invited to participate. If they agreed, “informed consent” was obtained and a “clinico-epidemiological data form” was

completed. On the day of the surgery, before anaesthetic induction, blood was collected in tubes with and without EDTA (BD Vacutainer, Becton and Dickinson). These blood samples were processed within 5 hours according to the procedures detailed in section 2.9. Serum and plasma were extracted and stored at -80°C. Immediately after the surgical resection, the specimen was opened through the intestinal wall opposite to the tumour. Fragments from the tumour and from the apparently normal adjacent mucosa (at least 10cm from the tumour) were collected in 10% formalin in phosphate buffered saline (PBS) pH 7.4, incubated for 24 hours, and then processed into paraffin blocks for future histological work. In order to obtain tissues for RNA expression analysis, we also collected tumour and adjacent samples in cryovials, and these were immediately frozen in liquid nitrogen. When this research project started, the investigator had already collected 50 pairs of samples (tumour and adjacent). However, although a specialised company was hired to transfer the material from Brazil to the UK, all samples defrosted during the shipment due to delays in the customs authority and insufficient dry ice in the package. After experiencing these problems, we decided to use a preservative solution (Allprotect®, Qiagen) instead of nitrogen snap-freezing to guarantee the preservation of tissues. Therefore, all the samples used in the mRNA expression analysis were placed immediately after collection in Allprotect®, incubated overnight at 4°C and stored at -80°C until they were used. Samples from cancer patients were collected between January 2013 and August 2015. After excluding samples which did not contain representative tissues and those which experienced technical problems during storage or transportation, the number of cancer samples suitable for analysis was:

i. Formalin-fixed, paraffin-embedded samples: 32 cases (tumour and adjacent mucosa)
ii. Allprotect®-treated tissue samples: 25 cases (tumour and adjacent mucosa).

b. Polyp-bearing individuals and normal controls

In order to obtain samples from individuals with colorectal polyps, we established a partnership with a colonoscopy clinic in Cuiaba (Dr Wladimyr Dias Moreno, Gastrocenter). Subjects referred to the clinic to undergo colonoscopy for a variety of symptoms (intestinal bleeding, constipation, diarrhoea, etc) or for CRC screening (asymptomatic) were identified, the research was explained and they were invited to participate. If they agreed, “informed consent” was obtained. During the colonoscopy, if one or more polyps were identified and no other alteration in the intestine was found, blood samples were collected and processed as explained in section 2.9. The polyps were then removed, immersed in formalin and sent for pathological analysis. This is the standard protocol after the identification of polyps and was not a research-related procedure. After the pathological assessment had been completed and the final report was released, the paraffin blocks were retrieved from the pathology laboratory for future histological and immunohistochemical evaluations. Alternatively, if no polyp, tumour or any other lesion was apparent, the individual was recruited as a normal control. Blood samples were collected and processed. Fragments from the normal mucosa were taken and placed into two flasks, one containing 10% buffered formalin (later processed into paraffin blocks) and the other with Allprotect® as explained above. Unfortunately, we were not able to collect fresh polyp samples either in liquid N_2 or in Allprotect® because the diagnostic routine demands that all polyps removed be sent for pathological analysis. The number of samples collected from these individuals was:

i. Polyp-bearing individuals: 18 adenoma samples.
ii. Normal controls: 10 intestinal mucosa samples.

c. Samples from cancer patients with more than 4 years follow-up for prognostic assessment - the prognostic cohort

To test whether or not a correlation existed between the expression of candidate biomarkers and clinical outcomes in CRC patients, we decided to perform an analysis of the expression of the proteins in a larger retrospective cohort. To do so, we established a partnership with a Brazilian pathologist (Dr Ivana Menezes) and a pathology laboratory in Cuiaba (Laboratorio Sao Nicolau). This is the reference laboratory for immunohistochemistry in that region. It has extensive expertise in this procedure and highly trained technical personnel.

The first phase of the collection was a survey in the records of two pathology laboratories (Laboratorio Sao Nicolau and the Julio Muller University Hospital Pathology Lab) aiming to identify cases of CRC who had undergone surgery more than four years ago. Next, we tracked the health service where the patient was initially evaluated to find whether the patient was followed up in that service or in any other service in which clinical information was traceable. Alternatively, if no clinical information was available, we checked the Brazilian electronic death database – Mortality Information System ("*Sistema de Informação de Mortalidade*" – *SIM*), a database with all deaths and its causes in the country. Lastly, we retrieved the paraffin blocks from these patients from the pathology lab archives. In total, we collected blocks from 96 patients. Twenty-one cases were excluded due to lack of tumour tissue in the block or because death occurred within 30 days from the operation, thus suggesting postoperative complication as the cause. Consequently, 75 cases were included in the final prognostic analysis (detailed in *Chapter 6*).

2.3. UK samples

Twenty-six individuals who underwent resection surgery at the Countess of Chester Hospital NHS Foundation Trust (Chester-UK) provided tissue samples that were included in this study. All specimens had malignant tumours from different stages, of which 7 also had normal adjacent mucosa in the block. Ten specimens had colonic polyps in addition to invasive tumours (6 low-grade and 4 high-grade adenomas). Lesions were classified as low- or high-grade dysplasia or invasive cancer according to the Royal College of Pathologists guidelines (Loughrey *et al*., 2014). Both the neoplasm and the adjacent mucosa (when present in the block) were included in a tissue microarray (TMA) and used for the analysis.

2.4. Clinicopathologic data collection

All Brazilian subjects involved in the prospective part of this research were interviewed by the investigators. During this interview, data regarding identification (name, name of the mother, address, phone number, national identification number, etc.), ethnicity and geographic background were collected. In addition, all information available concerning disease status such as scan or endoscopic test results was recorded. In patients who underwent a surgical procedure, the type of surgery and outcome was also added to the form. The final pathology report was obtained for every polyp and cancer patient, as well as the colonoscopy report for the polyp and control individuals.

Clinicopathologic meta-data from the UK cohort were retrospectively collected by Nadeem Al-Khafaji (University of Liverpool) using the MEDITECH and MediSecNet online databases, as part of his MRes research project.

Regarding the prognostic cohort, we assessed the records of the health service where the patient was followed up after surgery to obtain

survival information. In addition, the official Brazilian electronic death database was assessed in order to confirm that the mortality data were accurate. Overall survival was recorded as the interval between diagnosis and death from any cause (when death has occurred) or the date when the database was last checked (when death has not occurred).

2.5. Haematoxylin and eosin (H&E) staining and pathological review

In order to ascertain the presence of the target tissue in the paraffin blocks, each sample was evaluated by a pathologist either in Brazil (Dr Ivana Menezes, Sao Nicolau lab) or in the UK (Dr Timothy Andrews, Royal Liverpool University Hospital) using H&E staining. The staining was performed as follows: 4-micrometre sections were obtained from each block using a rotational manual microtome. Sections were dewaxed in xylene twice (5 minutes each), and then hydrated in a series of ethanol solutions with decreasing concentration (100% twice, 3 minutes each; 90% and 70%, two minutes each) and distilled de-ionised water (5 minutes) in a fume hood. Next, the sections were immersed in haematoxylin for 3 minutes and washed in running tap water during 10 minutes. Then, sections were incubated in eosin for 2.5 minutes and briefly washed in tap water. The stained sections were then dehydrated in increasing concentrations of ethanol (70%, 90% and 100% twice for 2 minutes each). Finally, sections were cleared in xylene twice (5 minutes each) and mounted. The same protocol was used in our laboratory every time we needed to re-evaluate the tissue present in the blocks (for example, after many sections had been taken from the block to reassure that target tissues were still present).

2.6. Immunohistochemistry (IHC)

2.6.1. Initial validation of candidate biomarkers

Samples collected from the Countess of Chester Hospital (UK-cohort) were arranged in a tissue microarray (TMA). A TMA is a cost-effective way of exploring the immunohistochemical expression of biomarkers in cancer tissues (Milanes-Yearsley *et al*., 2002, Ilyas *et al*., 2013). It is constructed by the transference of paraffin embedded tissue cores from several donor blocks into a single recipient block, therefore allowing the analysis of dozens or hundreds of samples in a single slide (Parsons and Grabsch, 2009). All individual paraffin blocks were sectioned, stained with H&E and evaluated by an experienced pathologist (Dr Timothy Andrews). The most representative areas were marked in the slide in order to facilitate tissue collection. Using a manual tissue microarrayer (MTA-1, Beecher Instruments Inc., Sun Prairie, Wisconsin, USA), three cores (0.6mm each) from each block were collected and transferred into recipient blocks. All subsequent analyses of the UK-cohort were carried out using TMA sections. The same procedure was performed using the Brazilian cohort for the construction of a TMA. However, we experienced a very high rate of core detachment from the slides produced using this microarray. The reasons for this problem were not clear, but it might involve the fixation process, the quality of the paraffin or even the conditions of storage of the blocks. Several procedures were attempted to decrease the rate of tissue loss (longer APES coating, incubation at 58°C prior to dewaxing, different slide brands, etc). Nonetheless, the loss of cores was still too high to allow the use of TMAs for the analysis of the Brazilian cohort. As a result, all the subsequent IHC staining procedures involving these samples were performed using individual whole sections.

Paraffin blocks (either TMAs or individual blocks) were cut into 4-micrometre sections. These sections were placed onto aminopropyltriethoxysilane (APES – Sigma, Gillingham, UK) -coated glass slides and left to dry in a laboratory oven at 37°C overnight. The slides were dewaxed in xylene as described for H&E. Then, the slides were immersed in

100% ethanol twice for 5 minutes. In order to block the activity of endogenous peroxidases, the slides were incubated in a solution containing 3% hydrogen peroxide in methanol for 10 minutes. Next, the slides were rehydrated in ethanol solutions (90% and 70%, 2 minutes each) and in distilled, de-ionised water (5 minutes). Afterwards, the sections were washed in Tris-buffered saline-Tween 0.1% (TBS-T) in an orbital shaker twice (5 minutes each). Heat-induced epitope retrieval was performed by immersing the slides in 10 mM citrate buffer pH 6.0. Heat was generated using a microwave oven at full power (800 watts) for two 10-minute sessions, with more citrate buffer added in the interval to top up the container and prevent the samples from drying out. The slides were then allowed to cool down for 10 minutes and washed in running tap water for additional 10 minutes. After that, the sections were washed in Tris-buffered saline (TBS) three times (from this step on, all TBS washes were performed for 5 minutes in an orbital shaker). To prevent secondary antibodies from binding to non-specific epitopes, the sections were incubated with 10% normal goat serum (Dako, Ely, UK) in TBS-T for 45 minutes at room temperature. Then, the slides were incubated with the primary antibody diluted in 10% normal goat serum in TBS-T overnight, at 4°C, in a humid chamber, using specific and optimised dilutions for each antibody tested (see table 2.1 for antibodies specifications and table 2.2 for reagent specifications). Next day, the sections were washed twice in TBS and a biotinylated secondary antibody solution was applied; incubation occurred during 30 minutes at room temperature. Goat anti-mouse or goat anti-rabbit secondary antibodies were used according to the species in which the primary antibody was raised, at a dilution of 1:200 in 5% normal goat serum in TBS. Slides were again washed twice in TBS and incubated with an avidin-biotin-peroxidase complex (Vectastain Elite ABC kit – Peterborough, UK) for 30 minutes at room temperature. After two more washes in TBS, the stain development was performed by applying a pre-prepared 3,3'-diaminobenzidine (Sigmafast DAB tablets – Sigma, Gillingham, UK) substrate solution onto the slides, followed by incubation for 4-5 minutes at room temperature. When satisfactory staining had been achieved, further colour development was blocked by briefly washing the slides with distilled de-ionised water and by a wash in TBS for 5 minutes. Finally, sections were

counterstained with haematoxylin for 5 minutes and washed in running tap water until the desired blue intensity had been achieved (usually 5 to 10 minutes). Then, the slides were dehydrated in increasing concentrations of ethanol (70% and 90%, 2 minutes each and 100% twice for 3 minutes each) and cleared in xylene twice (5 minutes each) before being mounted using DPX™ mounting medium (Sigma) and glass coverslips.

Table 2.1. Antibodies and specifications used in the IHC experiments.

PRIMARY ANTIBODIES		
Antibody/manufacturer	**Host/clonality**	**Dilution in 10% goat serum in TBS-T**
β-catenin/BD transduction laboratories	Mouse/polyclonal	1:50
NAP1L1/Abcam	Rabbit/polyclonal	1:4,000
RPL6/Proteintech	Rabbit/polyclonal	1:200
Prohibitin/Abcam	Rabbit/monoclonal	1:200
HMGB1/Abcam	Rabbit/polyclonal	1:1,000
SC35 or SFRS2/Abcam	Mouse/monoclonal	1:250
CDC5L/Abcam	Rabbit/monoclonal	1:10,000
SECONDARY ANTIBODIES	**(biotinylated)**	
Antibody/manufacturer	**Host/clonality**	**Dilution in 5% goat serum in TBS**
Anti-rabbit/Dako	Goat/polyclonal	1:200
Anti-mouse/Dako	Goat/polyclonal	1:200

Table 2.2. Reagents and buffers used in the IHC experiments.

Reagent/buffers	**Composition/manufacturers**
TBS	6.05 g of TRIS base (Sigma) and 8.76 g of NaCl (Sigma) to 1 litre of distilled water and adjusting pH to 7.4.
TBS-T	1mL of Tween 20 (Sigma) added to 1 litre of TBS.
Citrate buffer	2.94 g of Tri-sodium citrate dihydrate (Sigma) to 1 litre of distilled water and adjusting pH to 6.
H_2O_2-methanol	12 mL of 30% H_2O_2 (Sigma) in 400 mL of methanol.
Avidin-biotin-horseradish peroxidase complex	2 drops of reagent A and 2 drops of reagent B (Vectastain Elite ABC kit) in 5 mL of TBS, prepared 30 minutes before use.
DAB	1 tablet of Sigmafast DAB (Sigma) plus 5µL of 30% H_2O_2 (Sigma) in 5 mL of TBS (protected from light).

IHC stained slides were photographed for subsequent analysis. Whole-sections were recorded at 200x and 400x magnification using a Leica laser microdissection microscope and camera set (Leica Biosystems, Milton Keynes, UK). TMA slides were scanned at 200x magnification using the Aperio SlideScanner platform (Leica Biosystems, Milton Keynes, UK). The protocols used for scoring the images are detailed in the chapter describing the IHC validation work (*Chapter 3*).

2.6.2. **Prognostic study**

For the assessment of the prognostic significance of the expression of the candidate biomarkers (*Chapter 6*), a different IHC protocol was used. Most clinical pathology laboratories currently use pre-optimised buffers and solutions in order to save time and increase productivity. Therefore, the validation of our candidate biomarkers using techniques similar to the standard practices of clinical labs would increase the external validity of any findings. The protocol adopted in this part of the research was the routine

technique used in the Sao Nicolau laboratory (Cuiaba/Brazil), a pathology lab which has extensive expertise in IHC procedures. The experiments are described below (all branded solutions and buffers were purchased from Cell-Marque™/Sigma-Aldrich [Rocklin, California, USA]): Four micrometre sections were cut using a rotary microtome. Slides were dried at 60°C for 20 minutes. Then, samples were dewaxed in xylene and rehydrated as described before. After a wash step in distilled water, slides were immersed in Trilogy™ pre-treatment solution and incubated at 96°C for 30 minutes for epitope retrieval. After this, the slides were allowed to cool down and were washed in phosphate buffered saline (PBS) for 10 minutes. Peroxide block™ solution was then added and samples were incubated for 20 minutes. Another PBS wash was performed for 5 minutes. Next, the primary antibody solution (same concentrations as those described above) was placed onto the samples and incubation for 20 minutes was allowed at room temperature. Slides were washed in PBS for 5 minutes. HiDef Detection™ amplifier (secondary antibody solution) was then applied to the slides and they were incubated for 15 minutes. After a 5-minute PBS wash, the former step was repeated using HiDef Detection™ detector (a horseradish peroxidase polymer solution). Again, the slides were washed in PBS for 5 minutes. Finally, colour development was performed by incubating the slides with DAB substrate™ chromogen for 3 minutes. Stained slides were counterstained with haematoxylin, dehydrated and mounted as described earlier. Lastly, slides were photographed at 200x and 400x magnification using an Axio Scope.A1 microscope coupled with an AxioCam HRc camera (Zeiss, Oberkochen, Germany). These digital images were used for electronic scoring. Noteworthy, some samples were stained using both the protocol described here and that in the previous section in order to confirm that the staining pattern was similar. These comparative staining procedures were performed at different time points during the research and the results are discussed in the respective chapters (*Chapters 3* and *6*).

2.7. Quantitative-polymerase chain reaction (qPCR)

RNA extraction from tissues

Each tissue sample was taken from the -80°C freezer and defrosted on ice. Fragments of 20 to 30 µg of tumour or adjacent tissue samples were used, as recommended by the manufacturer of the RNA extraction kit (RNeasy® Mini Kit, Qiagen, Hilden, Germany). Regarding normal control samples, as these were collected by colonoscopy-guided forceps biopsies which yield very small fragments, we used all the material available for each case, usually 7 to 10 µg of tissue. Samples were placed into 1.5mL tubes containing 600µL of buffer RLT (proprietary solution rich in guanidine isothiocyanate) and were disrupted using a rotor-stator homogeniser (TissueRuptor®, Qiagen, Hilden, Germany) using short pulses of 30 seconds to avoid overheating the sample. The lysate was centrifuged at full speed for 3 minutes. The supernatant was carefully removed by pipetting and this was transferred into a new microcentrifuge tube. One volume of 70% ethanol was added and the solution was gently mixed by pipetting. Up to 700µL of this sample was then transferred to an RNeasy spin column, which was placed into a 2mL collection tube and centrifuged at ≥ 8000x g for 15 seconds. The flow-through was discarded and on-column DNase treatment was performed as follows: 350µL of buffer RW1 (proprietary ethanol solution rich in guanidine salt) was added to the column, which was centrifuged at ≥ 8000x g for 15 seconds and the flow-through was discarded. The DNase incubation mix was prepared by adding 10µL of DNase I stock solution and 70µL of buffer RDD, a proprietary solution to provide conditions for efficient on-column DNA digestion (both reagents provided in the RNase-free DNase Set, Qiagen, Hilden, Germany). This solution was placed onto the column and left on the bench top for 15 minutes. After this, 350µL of buffer RW1 was added to the column, centrifuged at ≥ 8000x g for 15 seconds and the flow-through was discarded. After DNase treatment, 500µL of buffer RPE (proprietary washing buffer containing ethanol) was added to the column and centrifuged at ≥ 8000x g for 15 seconds. This step was repeated with the same volume

of buffer RPE and centrifuged at ≥ 8000x g for 2 minutes in order to dry out the column membrane and remove any residual ethanol. Finally, the spin column was placed in another clean and labelled tube and 30µL of RNase-free water was added to the column to elute the RNA. The column/tube set was centrifuged at ≥ 8000x g for 1 minute to collect the eluted RNA. This step was repeated using the collected eluate in order to increase RNA concentration. Extracted RNA samples were stored at -80°C until use.

RNA extraction from cell lines

The extraction of RNA from cells lines was also performed using the RNeasy® Mini Kit. Procedures were essentially the same as described for tissue samples, except the initial tissue disruption step. Instead, cells were harvested by trypsinisation. Then, approximately 2.0 x 10^6 cells were centrifuged at 300x *g* for 5 min and the supernatant was discarded. Next, 350µL of buffer RLT were added to the pellet, vortexed and homogenised using a rotor-stator. Subsequently, 350µL 70% ethanol was added to the homogenate and mixed by pipetting. Thereafter, RNA extraction procedures followed the same steps described previously.

Assessment of RNA quality

RNA quality was assessed in terms of quantity, purity and integrity prior to any downstream applications. Quantity and purity were analysed by measuring the absorbance of the RNA solution using a NanoDrop 2000® spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Two microlitres of each sample were placed in the optical pedestal and the absorbance at 260nm and 280nm was measured. An absorbance of 1.0 at 260nm is equivalent to approximately 40 ng/µL of RNA. An A260/280 ratio between 1.8 and 2.2 indicates “pure” RNA. In this study, only samples with more than 50 ng/µL RNA concentration and with A260/280 readings between 1.9 and 2.1 were used. RNA integrity was assessed by the use of a

denaturing “bleach” agarose gel electrophoresis according to a published protocol (Aranda *et al*., 2012). The presence of bands corresponding to 28S and 18S ribosomal RNAs was assessed. Only samples with distinct 28S and 18S bands in an intensity ratio of approximately 2:1 were used.

First strand cDNA synthesis

After RNA extraction, the next step was the synthesis of the complementary DNA (cDNA). For this purpose, we used the TaqMan® Reverse Transcription Reagents kit (Applied Biosystems, Carlsbad, CA, USA) according to the manufacturer´s protocol, as below:

For each cDNA synthesis reaction (20 µL final volume), we added 2.0µL of 10x RT buffer, 1.4µL of 25mM $MgCl_2$, 4.0µL of 10mM dNTP mix, 1.0µL of RNase inhibitor (20U/µL), 1.0µL of MultiScribe™ Reverse Transcriptase (50U/µL), 1.0µL of 50 µM Oligo $d(T)_{16}$ and the RNA template. Nuclease-free water was added to make up to the final volume. Although the manufacturer states that up to 1µg of RNA can be used per reaction, excessive amounts of extracted RNA can cause problems. The presence of RT inhibitors mainly due to the carry-over of contaminants or internal inhibitors can affect RT efficiency (Pugniere *et al*., 2011, Bustin *et al*., 2015). Quantities of RNA less than the maximum 1µg may render some RT reactions more efficient. For this reasons, it is suggested that each researcher test the efficiency of the RT kit with different amounts of RNA input, a recommendation supported by the manufacturer of the kit used in this research (“*Relative Gene Expression Workflow document”*, available at http://tools.thermofisher.com/content/sfs/brochures/cms_075428.pdf, last accessed in 29/12/15). Therefore, we determined the reverse transcription dynamic range as per the protocol provided by the document above. Briefly, we carried out RT reactions using a serial dilution of RNA input, from 1µg to 62.5ng (i.e. from 1/1 to 1/16 of the maximum recommended quantity). The cDNA generated from these diluted samples was then assessed by qPCR and standard curves were produced so that the efficiency of the RT reactions

using different RNA input was verified. These procedures were performed for all the biomarkers that we intended to test. These experiments established that RNA input of approximately 300ng was the appropriate amount for cDNA synthesis. Consequently, all cDNA produced from human samples was derived from reactions of 20µL using 300ng of RNA as template.

qPCR equipment and reagents

Quantitative PCR reactions were performed using the 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). All reactions were carried out in triplicate. Non-template controls (no cDNA added) were used in every qPCR run. For each reaction, the following reagents were used: 10µL of TaqMan® Fast Advanced Master Mix; 1µL of TaqMan® Gene Expression Assay (detailed below) for the target; 1µL of TaqMan® Gene Expression Assay for the endogenous control (*ACTB*) as a duplex reaction; 2µL of diluted cDNA template and 6µL of nuclease-free water to complete the final volume of the reaction. Given the extreme sensitivity of qPCR, many researchers recommend that the cDNA produced during the reverse transcription be diluted before using it for gene expression assessment. This can decrease the presence of PCR inhibitors from the RT step in the same way explained earlier and allow the use of minor quantities of precious samples per reaction. However, overdilution can cause a given sample to reach the exponential phase of amplification too late during the PCR (after cycle 35, for example), thus impairing the accuracy of the method. To define the optimal cDNA dilution for our experiments, we tested different concentrations of cDNA samples using assays for all targets in order to make sure the threshold cycle (Ct) was between cycles 15 and 30. In these experiments, we used cDNA in the following dilutions: 1:1, 1:2, 1:4, 1:8 and 1:16. The results of this evaluation demonstrated that all dilution factors were suitable for the analysis (see *Chapter 4*), and we chose the dilution factor 1:4 to be used in subsequent qPCR experiments. TaqMan® Gene Expression Assays were used in this work. These “single-tube assays” contain a set of primers (reverse and forward), a fluorescent probe specific to the gene

sequence being studied and an optimised buffer. The specifications of the assays used are described in the table below:

Table 2.3. Specifications of the TaqMan® gene expression assays used in this experiment. Beta-actin was used as the reference gene (or internal control).

Gene Symbol	Entrez Gene ID	Description	NCBI Location Chromosome	TaqMan® assay Id(*)
CTNNB1	1499	catenin (cadherin-associated protein), beta 1	Chr.3: 41240942 - 41281939	Hs00355049_m1
NAP1L1	4673	nucleosome assembly protein 1-like 1	Chr.12: 76438672 - 76478738	Hs00748775_s1
RPL6	6128	ribosomal protein L6	Chr.12: 112842994 - 112847443	Hs03044365_g1
PHB	5245	prohibitin	Chr.17: 47481420 - 47492242	Hs00855044_g1
ACTB	60	actin, beta	Chr.7: 5566779 - 5570232	Hs99999903_m1

* Primer and probe sequences for pre-optimised TaqMan® assays are patent-protected. Therefore, the assay Id is mentioned.

Cycling conditions were as follows: 1 initial hold for 2 minutes at 50ºC for uracil-N glycosylase (UNG) incubation; 1 hold for 20 seconds at 95ºC (polymerase activation), 40 PCR cycles of 3 seconds at 95°C (denaturation) followed by 30 seconds at 60°C (annealing/extension). These conditions are specific for the use of TaqMan® Fast Advanced Master Mix.

Data analysis

Results were produced using the method of comparative CT, also referred to as the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001, Schmittgen

and Livak, 2008). This is a relative quantification technique in which the difference in the expression of the gene of interest in different conditions is assessed relative to some housekeeping gene (also known as endogenous control or reference gene). Details of the procedures involved in the qPCR analyses are presented in the dedicated chapter along with the results and discussion (*Chapter 4*).

2.8. Cell culture and RNA interference experiments

Cell line

CRC cell lines have been widely used in studies assessing tumour biology, drug response and biomarkers (Mouradov *et al*., 2014). HCT116 cells are derived from a colonic adenocarcinoma removed from a male patient (Ilyas *et al*., 1997, Morin *et al*., 1997). In this study, both *TP53* wild-type and *TP53* null HCT116 cells were used (Bunz *et al*., 1998). Cell lines were provided by Professor Bert Vogelstein (Johns Hopkins University, USA).

Culture reagents and protocol

The reagents and conditions used for cell culture are described below:

Growth medium: McCoy´s 5a + Glutamax™ medium (Gibco®) + 10% heat-inactivated foetal bovine serum (Gibco®) + 100 units/ml penicillin + 100µg/ml streptomycin (Sigma®).

Protocol:

A vial containing 1 x 10^6 cells was taken out of the liquid nitrogen and thawed in a 37°C water bath. Cells were then re-suspended in 15mL of growth medium and transferred into a 75 cm^2 flask. The flask was placed into

a 5% CO_2 incubator at 37°C. Next day, the growth medium was replaced to avoid toxicity caused by the DMSO contained in the freezing medium. When 70-90% confluence was achieved, growth medium was aspirated and discarded. The cell layer was briefly rinsed with 5mL of PBS (Gibco®) and cells were detached from the flask by trypsinisation using 5mL of 0.25% trypsin-EDTA (Sigma®). When the cells were dispersed, 5mL of complete growth medium was added and the cells were collected by gentle pipetting. Then, 1mL of cell suspension was added to a new flask and completed with growth medium to 15mL. Flasks were placed into a 5% CO_2 incubator at 37°C. Growth medium was changed every 48 hours if split confluence had not been reached.

Transfection reagents and protocol

Transfection conditions previously optimised in our laboratory were used in these experiments. All transfection reagents were purchased from GE Dharmacon™ (Lafayette, CO, USA). Specific siRNAs (ON-TARGETplus SMART pool siRNA®) targeting *NAP1L1* (product number L-017274-01-0005) and *RPL6* (product number L-012955-00-0005) were used. Additionally, a *GAPDH*-directed siRNA pool (product number L-001830-10-05) and a non-targeting (scrambled) siRNA pool (product number L-001810-10-05) were also used as positive and negative controls, respectively. The delivery of siRNA to cells is impaired by its negative charge which hinders cell membrane penetration. In addition, the presence of RNAses in the culture medium also makes the use of siRNAs challenging. To circumvent these issues, methods of nucleic acid delivery have been developed such as viral vectors, electroporation and liposomal delivery (Gao and Huang, 2009). GE Dharmacon™ offers liposomal-based transfection reagents (TRs) optimised for siRNA experiments using different cell lines. The company suggests Dharmafect 2® (hereafter referred to as TR2) as the ideal medium for transfecting HCT116 cells. Therefore, this was the selected option.

Using the same reagents described in the “Culture reagents and protocol” section above, we prepared distinct culture media to be used in the transfection procedures, as below:

- Complete medium: same components described as “growth medium” above, used for maintaining cells before transfection;
- Antibiotic-free medium: same components of “complete medium” except the antibiotics. This medium was used 24 hours before, during and after transfection, as cell membranes become very permeable to various substances, including antibiotics, upon the use of transfection reagents. This may cause excessive cell toxicity and should be avoided;
- Serum-free medium (glutamine only): used for preparing and diluting the siRNAs before adding the solution to the antibiotic-free medium.

The protocol used for siRNA transfection is summarised below:

a. Initially, lyophilised siRNA pools were resuspended in 1x siRNA buffer at 20µM stock concentration and analysed in a Nanodrop™ to ascertain the RNA concentration. This solution may be kept at 4°C for up to 6 weeks or aliquoted and frozen (-20°C);
b. Cells reaching 70-80% confluence were trypsinised, detached and counted using a Bio-Rad TC10™ Automated Cell Counter;
c. Using 6-well culture plates, 300,000 cells per well were allowed to attach for 24 hours in 2mL of antibiotic-free medium;
d. Next day, siRNA reagent was prepared by diluting the 20µM stock solution in 1x siRNA buffer to a concentration of 5µM. In separate tubes, working siRNA and TR2 solutions were prepared (volumes per well):

    Tube A: siRNA solution: 20µL of 5µM siRNA were added to 180µL of serum-free medium, gently mixed and incubated for 5 minutes;

Tube B: TR2 solution: 4µL of TR2 were added to 196µL of serum-free medium, gently mixed and incubated for 5 minutes;

e. The contents of tubes A and B were added, gently mixed and incubated for 20 minutes;
f. 1.6mL of antibiotic-free complete medium was added and gently mixed. Final transfection solution contained siRNA at 50nM;
g. Culture medium in plate wells was aspirated and discarded. Then, 2.0mL of the transfection solution was added to each well;
h. Cells were incubated in a 5% CO_2 incubator at 37°C;
i. Cells were harvested by trypsinisation, counted as above and kept on ice until final use.

All the transfection procedures were carried out using *NAP1L1*, *RPL6*, *GAPDH* and non-targeting siRNAs, and included both *TP53* wild-type and null HCT116 cells. Additionally, non-transfected cells were analysed alongside to assess the toxicity caused by the transfection substances. Experiments were run in duplicate for each condition. Quantitative PCR was used to confirm knockdown efficiency. From reagent preparation to cell handling, all procedures were performed in sterile conditions within tissue culture hoods.

Sulforhodamine B proliferation assay

Sulforhodamine B (SRB) assay is a widely used proliferation test based on the estimation of cellular density via the determination of total protein content in culture plate wells (Vichai and Kirtikara, 2006). In this study, HCT116 cells (*TP53* wild-type and null) underwent siRNA experiments as described above. After 48 hours of incubation with the transfection reagents, cells were harvested, counted and used in SRB assays as described below (Vichai and Kirtikara, 2006):

a. Cell suspension concentrations were adjusted to 1,000 cells per 100µL using complete medium;

b. Using 96-well plates, 100µL of cell suspension from each condition was added to plates in 6 replicate wells. The format was set as untransfected, non-targeting siRNA and targeting siRNA in adjacent columns. *NAP1L1*- and *RPL6*-silenced cells were assessed in separate plates. Three plates were used for each protein in order to assess SRB staining 24, 48 and 72 hours after plating;
c. Plates were incubated at 37°C in 5% CO_2. At the time-points above mentioned, a plate for each protein was collected and fixed, as below:
    - Without removing the medium, 100µL of cold 10% trichloroacetic acid (Sigma) was added to each well and incubated at 4ºC for 1 hour;
    - Plates were then washed in slow-running tap water. Excess water was removed using paper towels. Plates were allowed to air dry overnight and were stored until the last plates were fixed;
d. When all plates had been fixed and dried, SRB staining was performed:
    - 100µL of 0.4% SRB (Sigma) in 1% acetic acid (Fisher Scientific UK) was added to each well and incubated at room temperature for 30 minutes;
    - Plates were rinsed 4 times with 1% acetic acid solution. Excess liquid was removed using paper towels;
e. A basic solution (200µL of 10mM Tris base solution, pH 10.5, Sigma) was added to each well and incubated in an orbital shaker for 5 minutes;
f. Optical density was measured at 510nm using a Tecan Sunrise™ 96-well plate reader. Signal-to-background ratios were calculated.

Gene expression array plates

In order to explore possible effects of our biomarkers on other CRC-related genes and pathways, we decide to test how gene silencing via siRNA

targeting our candidates would affect the expression of a collection of genes known to be involved in CRC development. To achieve this, gene candidates whose silencing resulted in impaired cell proliferation were identified in the siRNA experiments described above. Then, RNA was extracted and cDNA was synthesised according to the protocols explained in section 2.6. These cDNA samples were loaded into Human Developmental Phases of Colorectal Cancer® array plates (Applied Biosystems, Carlsbad, CA, USA) and run in a 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA), using the cycling conditions recommended by the plate manual. Details of these experiments, including the genes assessed, are described in the corresponding sub-chapter (*section 4.6.3*).

2.9. ELISA

Sample collection, processing and storage

Blood samples were collected from healthy controls, individuals with colorectal adenoma or colorectal cancer patients and processed no more than 5 hours after collection. For serum extraction, blood was collected in anticoagulant-free Vacutainer tubes (Becton Dickinson). Tubes were left undisturbed at room temperature for at least 30 minutes until a firm clot had been formed. Next, each tube was centrifuged at 2,000x g for 10 minutes at 4ºC. The supernatant (serum) was transferred into cryovials and was stored in -80°C freezers until use or shipping. For plasma extraction, EDTA-treated Vacutainer tubes were used. Tubes were centrifuged for 10 minutes at 2,000x g. The supernatant (plasma) was carefully transferred into cryovials and stored as describe for serum. Samples collected in Brazil were transported to the UK by a company that specialised in biological sample transportation (Biocair International Ltd) and were kept under freezing temperatures during the entire shipment time by the use of dry ice. Upon arrival at the destination, samples were immediately placed in scientific freezers at -80°C. The source and number of samples used in this research are detailed in the dedicated chapter covering immunoassays (*Chapter 5*).

Assay procedure

On the day of the first experiment, samples were defrosted and kept on ice. Each sample was aliquoted into several cryovials, each one containing 120µl of serum or plasma. The aliquots to be used immediately were diluted at 1:3 using sterile PBS, and were used in the ELISA experiments. All remaining aliquots were placed in the -80°C freezer for later use. In subsequent experiments, individual vials were defrosted, diluted as described and totally used (any leftover was discarded), thus avoiding several freeze-thaw cycles.

Although we tested various ELISA kits from different manufacturers, they all shared almost identical assay protocols. General procedures for the ELISA experiments are summarised below (all solutions provided with the kits):

1. Prepare all reagents, samples and standards;
2. Add 100µl of standards or samples to each well. Incubate for 2 hours at 37°C;
3. Aspirate and add 100µl of prepared Detection Reagent A. Incubate for 1 hour at 37°C;
4. Aspirate and wash 3 times;
5. Add 100µl of prepared Detection Reagent B. Incubate for 30 minutes at 37°C;
6. Aspirate and wash 5 times;
7. Add 90µl of Substrate Solution. Incubate for 15-25 minutes at 37°C (do not exceed 30 min);
8. Add 50µl of Stop Solution. Read plate at 450nm.

ELISA plates were read using a Tecan Sunrise™ 96-well plate reader at 450nm. Results were produced by loading plate results (raw readings) into an online ELISA results tool (www.elisaanalysis.com). The chosen method of analysis was the four-parameter logistic (4PL) regression model.

2.10. MSD-based electrochemiluminescence

Due to poor inter-plate reproducibility and low sensitivity demonstrated by the NAP1L1 ELISA kits used, the development of an in-house assay was added to the aims of this research project. The platform chosen was electrochemiluminescence using equipments and reagents provided by Meso Scale Discovery® Inc (MSD). This platform allows the development of immune-based assays with high sensitivity and specificity, and a wide dynamic range. All the procedures performed and the results produced are presented in a dedicated sub-chapter within this thesis (*Chapter 5, section 5.5*).

2.11. Statistical analysis

Several types of data were produced during this research. Continuous numerical variables (intervals or ratios) were first assessed regarding normality (Kolmogorov-Smirnov test) and homogeneity of variance (Levene´s test). When the parametric assumption was met, comparisons between groups were performed by using Student´s t-tests (two groups) or analysis of variance – ANOVA (three or more groups). Tukey´s HSD test (for equal variances) or Dunnett´s T3 test (for unequal variances) was used for pair-wise comparisons in this case. Alternatively, when normality could not be ascertained, non-parametric tests were used: Mann-Whitney U test (for comparing two groups) or Kruskal-Wallis test (for three or more groups). Dunn-Bonferroni test was the post-hoc comparative test of choice in the latter case. Correlation between continuous variables was assessed by using *Pearson´*s (for parametric data) or *Spearman´*s (for non-parametric data) correlation coefficients.

Categorical data were compared using the Chi-square test (or Fisher´s exact test in case of less than five expected counts per cell in the contingency table).

For the survival analysis, groups were assessed using the Kaplan-Meier method, and survival curves were compared by log-rank tests. When significant differences were observed, Cox proportional hazards model was used for multivariate analysis.

Two-sided p values <0.05 were accepted as significant in the entire study. All statistical analyses were performed using the software IBM® SPSS® Statistics version 22.

Chapter Three: Immunohistochemical validation of the candidate biomarkers in human tissues

3. CHAPTER 3 – IMMUNOHISTOCHEMICAL VALIDATION OF THE CANDIDATE BIOMARKERS IN HUMAN TISSUES

3.1. Introduction

Previous research conducted in our group using animal models of CRC based on *Apc* inactivation found potential biomarkers that were differentially expressed in tissues and blood from affected mice (Ibrahim, 2014, Hammoudi *et al*., 2013). If also present in human CRC samples, these alterations could potentially be useful in clinical practice as diagnostic tools. Furthermore, a biomarker may be used as a prognostic or predictive test, when it provides information regarding clinical outcome or response to specific treatments, respectively (Pritzker, 2015). Despite extensive research, few cancer biomarkers have hitherto progressed from basic research to the clinic (Diamandis, 2014). In this context, proteomic approaches such as the methods used in our previous work are amongst the most promising strategies for biomarker discovery (Li and Chan, 2014, Tjalsma, 2010, Alvarez-Chaver *et al*., 2014, de Wit *et al*., 2013).

Immunohistochemistry has recently become a widespread ancillary technique in histopathology (Lin and Chen, 2014). It is an inexpensive, reproducible, and widely available method for assessing protein content and localisation in cells and tissues. The technique has played a very important role in the identification of diagnostic, prognostic or predictive markers in many types of cancer (Chamberlain *et al*., 2015, Zaha, 2014, Toffart *et al*., 2014, Varma and Jasani, 2005). In CRC patients, IHC has mainly been used for assessing the status of various mismatch repair proteins (MLH1, MSH2 and MSH6) either as a screening procedure for hereditary non-polyposis colorectal cancer (HNPCC) (Steinhagen *et al*., 2012) or for prognostic stratification (Yoon *et al*., 2011). Several proteins have been tested in CRC tissues for their diagnostic or prognostic significance, although none has become a standard biomarker.

The aim of this chapter is to assess the immunohistochemical expression of selected candidate biomarkers in tissues from individuals with colorectal adenoma, cancer and normal controls. Additionally, in order to validate the candidates in different populations (with different ethnicity and genetic background), we used samples collected from health services in Brazil and in the UK, as detailed in *Chapter 2 - Methods*. For this purpose, we tested different scoring systems to assure accuracy, reproducibility and practicality. The process of selection and optimisation of an appropriate scoring method is described below.

3.2. Scoring systems: development, testing and optimisation

Despite being a powerful research and clinical technique, IHC has several potential pitfalls. Its performance is affected by many conditions, from sample collection and fixation to the final analytical steps (Lin and Chen, 2014). In addition, a standardised scoring system must, ideally, be used to increase objectivity and to assure the reproducibility of the method (Taylor, 2014). Unfortunately, none of the biomarkers evaluated in this project has an accepted scoring method. Among these proteins, β-catenin has been the most assessed in immunohistochemical studies. It has been evaluated both in colorectal cancer (Chen *et al*., 2013, Morikawa *et al*., 2011, Bruun *et al*., 2014) and in other neoplasms (Cuello-Carrion *et al*., 2015, Xu *et al*., 2015, Li *et al*., 2014a, Li *et al*., 2014b). However, there has been no agreement as to the best way of scoring tissues stained for β-catenin expression. Researchers have used a variety of quantitative or semi-quantitative methods for analysing the cytoplasmic or nuclear presence of this protein (Morikawa *et al*., 2011, Chen *et al*., 2013, Yoshida *et al*., 2015). Recent evidence suggests that the translocation of β-catenin into the nucleus is the critical event for colorectal carcinogenesis and that nuclear localisation is associated with tumour aggressiveness and poor survival (Yoshida *et al*., 2015, Chen *et al*., 2013). Therefore, for scoring β-catenin-stained animal tissues, our group has been using a system which takes into account the nuclear/cytoplasmic balance of staining (Ibrahim, 2014), as explained later.

Regarding the other biomarkers included in this study, most of them have been poorly evaluated by IHC to date or have not been evaluated at all (this will be further explored in the next sections). Our previous data suggested that most of these proteins also exhibited a nuclear/cytoplasmic imbalance in animal models of CRC. For that reason, in the absence of an established scoring method and taking into account the fact that these candidate biomarkers were derived from Wnt activated systems (as is the case for β-catenin itself), we decided to explore the nuclear/cytoplasmic localisation of these proteins as well.

We tested different means of scoring, from manual to electronic counting, and the results obtained using these different methods were compared in order to maximise objectivity and reproducibility. The process of development and optimisation of the scoring systems is presented below.

A modified H-score obtained by manual counting of individual cells

In pathology and research practice, most IHC slides are currently evaluated by a pathologist or investigator by visual inspection under a light microscope. In general, a scale-based scoring system is used to describe these observations. The first popular scoring protocol was the H-score which was proposed by McCarty *et al*. in 1986 (McCarty *et al*., 1986). This method is based on a semi-quantitative assessment of staining intensity and distribution, and uses a mathematical formula to produce a final score. Later, a simpler scoring system gained popularity: the Allred- or quick-score (Harvey *et al*., 1999). This procedure also assigns scores for the intensity and proportion of staining, but simply sums the values to produce a general score.

In this research, we started the work by scoring each slide using a method developed in our group for β-catenin-stained tissues, adapted from the H-score. It takes into account the intensity and localisation of the protein in the membrane, cytoplasm and nucleus of cells, and will hereafter be

referred to as a “modified H-score”. Score 0 represents a low level or negative staining of both the nucleus and the cytoplasm, with membranous staining only being observed; score 1 indicates increased staining of the cytoplasm, but little or no staining of the nucleus (cytoplasmic localisation); score 2 represents positive and equal staining of both the cytoplasm and the nucleus, and score 3 denotes strong nuclear staining, darker than that present in the cytoplasm (nuclear localisation). For this assessment, we used the scientific image manager ImageJ – an open-source software provided by the US National Institutes of Health that has been widely used for more than 25 years (Schneider *et al*., 2012) and which is publicly available via rsbweb.nih.gov/ij/ (last accessed in 26 September 2015). The “Cell Counter” plugin was used to individually score all identifiable target cells within the field (a minimum of 100 cells per field). After this, an overall score was generated using the formula: modified H-score = [(proportion of 0) x 0] + [(proportion of 1) x 1] + [(proportion of 2) x 2] + [(proportion of 3) x 3], with final scores ranging from 0 to 3. Figure 3.1 illustrates patterns of staining corresponding to different scores.

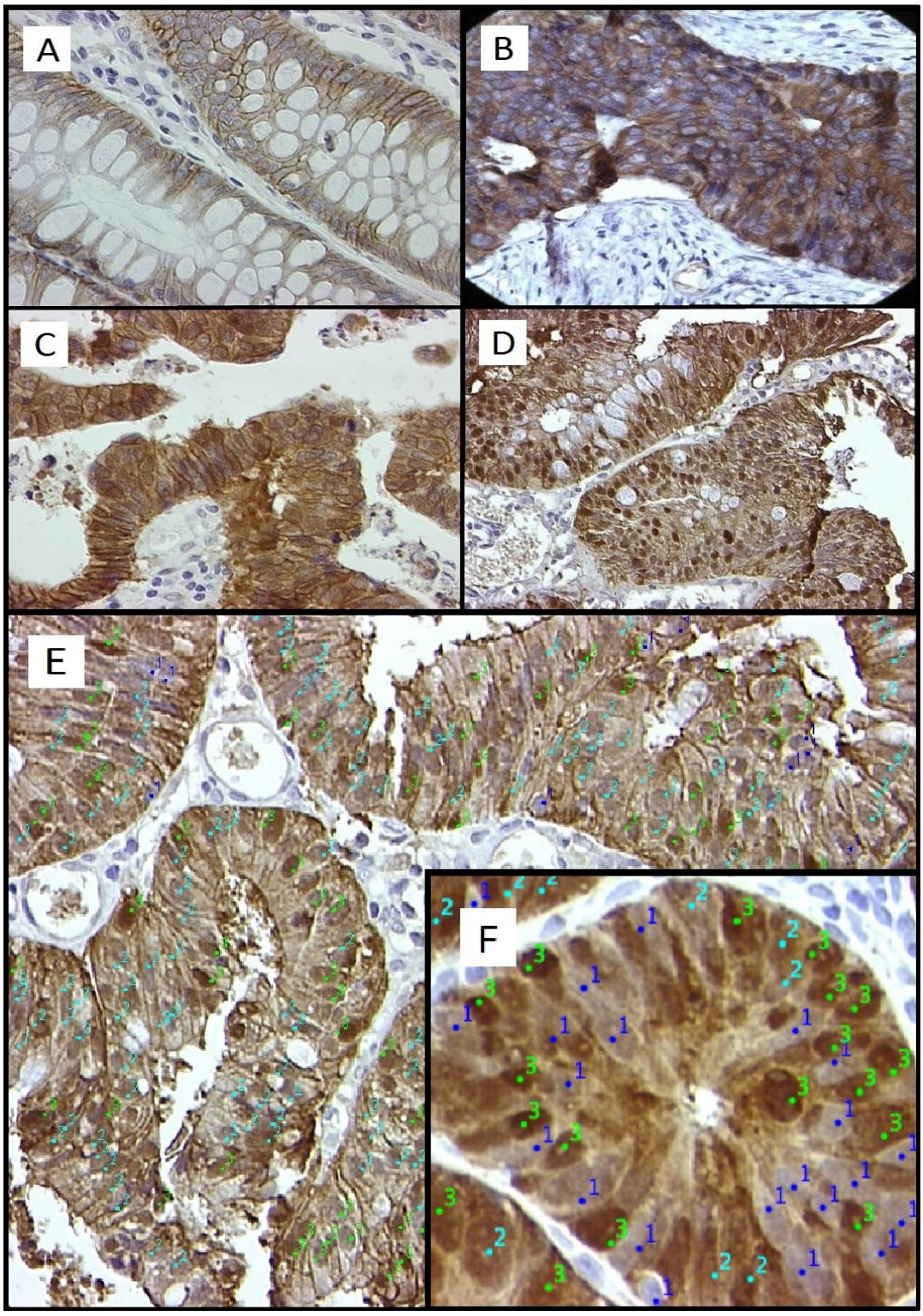
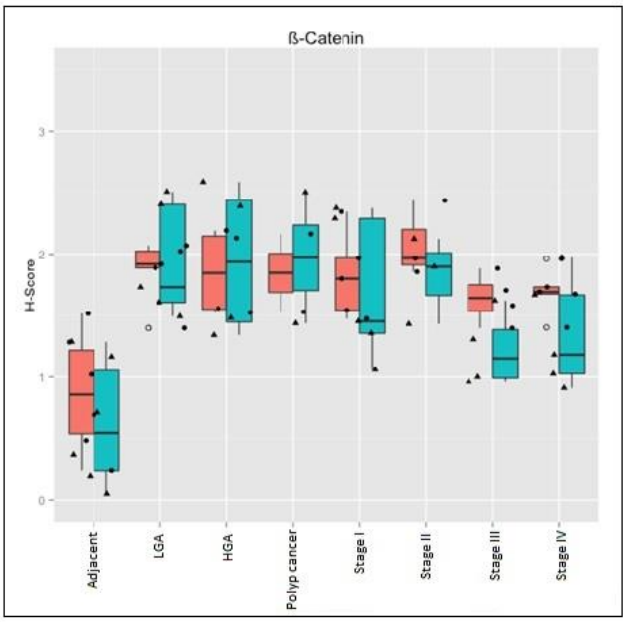
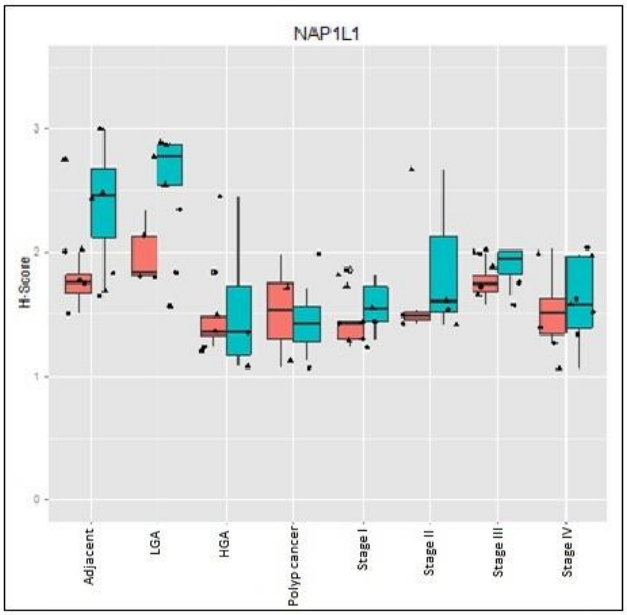


Figure 3.1. Modified H-score system explained. (A-D): tissues exhibiting relatively homogeneous staining patterns exemplifying modified H-scores of 0 (A, normal tissue), 1 (B, cancer), 2 (C, cancer) or 3 (D, cancer) (magnification: 200x). (E): tissue showing a much more common (heterogeneous) staining pattern (original magnification: 400x). In all cases, each individual cell in the field was assessed and assigned a score from 0 to 3 as seen in F (magnified image to show the scoring marks).

Using this method, we initially scored the UK-cohort of samples that had been stained for β-catenin, NAP1L1 and RPL6. In order to assess the reproducibility of the technique, two different researchers (Cleberson Queiroz (CQ) and Nadeem Al-Khafaji (NA)) performed the scoring separately. Both individuals are not trained pathologists. Therefore, the scoring process started after a brief period of training under the supervision of Dr Timothy Andrews (Liverpool – UK). During two days, the mentioned professional provided guidance for the correct identification of target areas in IHC slides. Tissue microarrays containing the samples were sectioned and stained according to the protocol described in *Chapter 2 - Methods*. After that, stained slides were scanned at 200x magnification using the Aperio SlideScanner platform (Leica Biosystems, Milton Keynes, UK). The modified H-score was calculated as explained above. Two fields were analysed for each sample. For this preliminary analysis, we divided the UK-cohort into the following groups, according to the pathologic diagnosis: normal adjacent colonic tissue, low-grade adenoma, high-grade adenoma, polyp cancer and cancer (subdivided into its different stages). As can be seen in figure 3.2, the final median scores produced by the two researchers exhibited similar patterns for the proteins tested. However, some discordant results were produced. Due to the fact that the trend was similar for most markers, a mean of the scores from both investigators was generated and this was used for the statistical analysis presented later. Consequently, the immunohistochemical analyses shown in the next sections for β-catenin, NAP1L1 and RPL6 relative to the UK-cohort were performed using the method described above. Unfortunately, several blocks retrieved from the Countess of Chester hospital, and that were used for the construction of the TMA, had insufficient target tissues to allow their use for the evaluation of the other candidates. As a result, the assessment of the remaining proteins was limited to analysis of the Brazilian cohort.





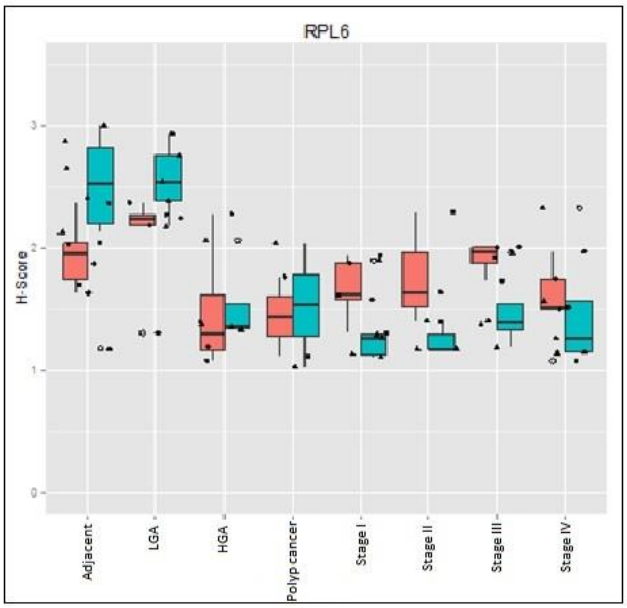


Figure 3.2. Comparison between the manual modified H-scores produced by two different researchers. On the x axis, the cases are divided into groups according to the pathologic diagnosis: normal adjacent colonic tissue; low-grade adenoma (LGA); high-grade adenoma (HGA); polyp cancer; invasive cancer (stages I, II, III and IV). Red boxes represent the scores from CQ, whereas green boxes indicate the scores from NA. The scatter-plot represents scores for individual cases. β-catenin scores from both researchers were highly consistent and similar, except for later stages. Regarding NAP1L1 and RPL6, the trend of modification in the scores from normal to cancer tissues is clearly the same according to the evaluation by both researchers. However, a difference in median scores is noted. Boxplots represent medians and interquartile ranges.

Analysing the nuclear and cytoplasmic expression of the biomarkers using electronic scoring tools

Despite being accepted scoring methods, H-score and quick-score (or any other procedures derived from them) have important and common pitfalls. They are time-consuming and, most importantly, are influenced by the visual perception of the observer (Varghese *et al*., 2014, Jaraj *et al*., 2009, Taylor, 2014). As a result, minor differences in colour development time during the IHC assay or in haematoxylin counterstaining can affect the final score. Additionally, a long and intensive training in histopathology is necessary to decrease the degree of subjectivity in this analysis, although in specific situations, even trained medical pathologists exhibit a large inter-observer variability when assessing IHC markers (Jaraj *et al*., 2009).

For all these reasons, electronic tools for the automatic scoring of digital IHC images have been developed (Rizzardi *et al*., 2012). Most scoring softwares use the colour deconvolution algorithm, in which the brown staining produced by DAB is separated from the blue counterstaining from haematoxylin resulting in two separate images (Helps *et al*., 2012, Ruifrok *et al*., 2003, Ruifrok and Johnston, 2001), although alternative methods based on statistical modelling have recently been reported (Shu *et al*., 2016). The majority of these modern systems are expensive, use dedicated software and hardware, and require extensive pre-testing optimisation. These features have prevented the widespread use of this technology (Varghese *et al*., 2014). Nonetheless, there are a few simpler, free-of-charge applications available which are intended to facilitate the process of IHC scoring. These packages use the same colour deconvolution algorithm and do not require any proprietary programme or hardware. Some of these systems have been tested in various research scenarios and have been shown to have similar or better performances than visual scoring (Varghese *et al*., 2014, Tuominen *et al*., 2010, Sysel *et al*., 2013, Prasad *et al*., 2011, Hammes *et al*., 2007).

Therefore, in order to improve the accuracy and objectivity of the assessment of our candidate biomarkers, we decided to test two electronic scoring systems using samples from the Brazilian cohort of patients. The

reason for testing two methods is because none of the available systems allows the measurement of both the nuclear and the cytoplasmic expression of proteins concurrently. Hence, we opted for testing these tools separately using different applications. For this purpose, we used two freely-available plugins developed for ImageJ: “IHC Profiler” and “ImmunoRatio”. The optimisation and testing process for each of these methods is explained below.

The use of “IHC Profiler” for the cytoplasmic assessment of IHC stained slides

In 2014, Varghese *et al*. published the results of the development and validation of an electronic plugin for automatic IHC scoring compatible with ImageJ – IHC Profiler (Varghese *et al*., 2014). The authors tested this tool in tissues stained for several markers in a large and diverse set of cancer samples obtained from the Human Protein Atlas database (http://www.proteinatlas.org). The plugin was based on the colour deconvolution process (as depicted in figure 3.3) and was shown to be highly accurate in the analysis of the images when compared with visual scoring by a pathologist.

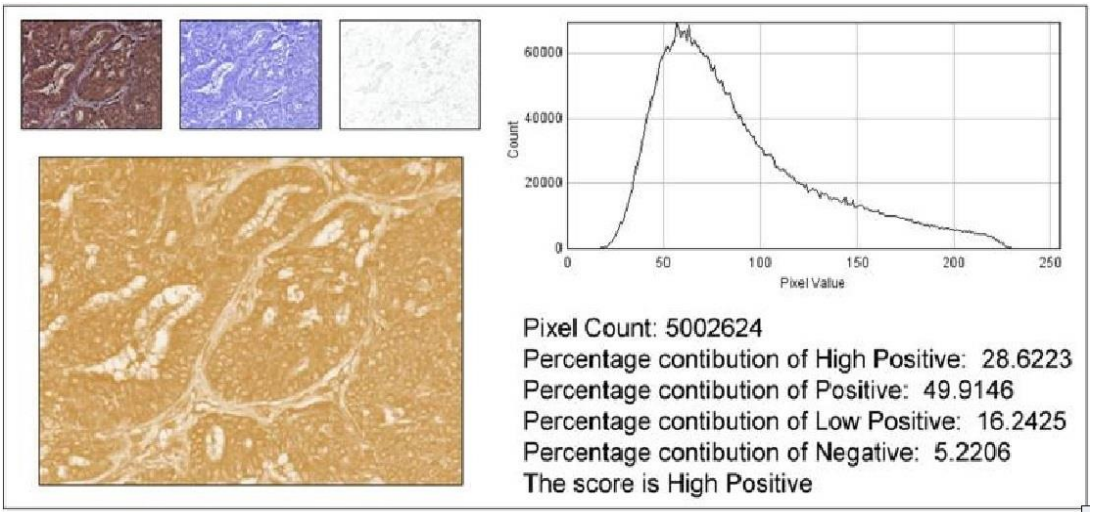


Figure 3.3. The colour deconvolution process used by the IHC Profiler plugin to generate separate images for DAB (brown) and haematoxylin (blue). Only the DAB-stained image is assessed for stain intensity. The result is presented as the proportions of areas with high-positive, positive, low-positive or negative staining and a final result is provided. From (Varghese *et al*., 2014).

In order to test the suitability of IHC Profiler for the assessment of the expression of our biomarkers, we initially analysed the expression of β-catenin in samples from the Brazilian cohort of patients. After staining according to the protocol previously described, each individual slide was analysed under the microscope and the images were recorded. Before actually analysing the image, we selected representative areas in the field of view in order to exclude as much non-target tissues (such as stroma, immune infiltrate, fat, staining artefacts, etc) as possible. This process is illustrated in figure 3.4.

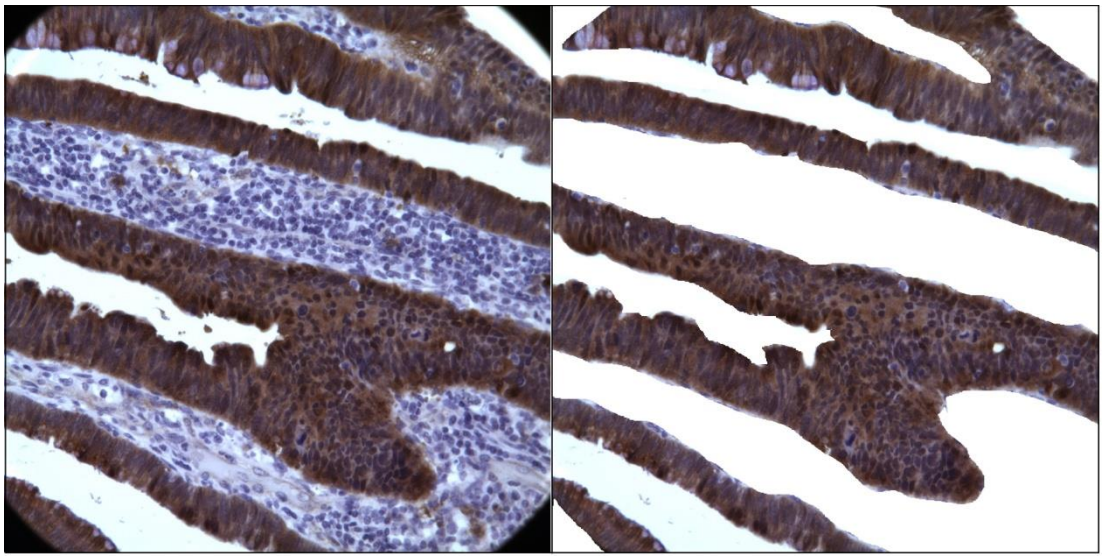


Figure 3.4. The selection of the target area to be analysed. As the plugins measure all stained areas in a field, images were edited in order to remove non-target tissues as shown in the picture. This procedure is important to reassure precision in the scoring process.

Next, we analysed all the samples using the cytoplasmic channel of IHC Profiler. This is a very straightforward process and involves selection of the image, identification and removal of non-target areas and application of the plugin. The result is presented as the percentage of areas with negative, low-positive, positive and high-positive scores. Additionally, an overall result is provided using one of these categories. However, this “final score” is less sensitive than using the proportions as tissues with small but significant differences in the staining pattern may be categorised in the same group, as can be noted in figure 3.5. Therefore, in order to improve accuracy and allow

the identification of minor differences in tissue staining, we decided to ignore this overall score and use the proportions of different staining intensities to generate an H-score-like value, using a formula similar to the one that we previously used for the modified H-score in the analysis of the UK-cohort: modified IHC Profiler score = [(proportion of negative) x 0] + [(proportion of low-positive) x 100] + [(proportion of positive) x 200] + [(proportion of high-positive) x 300], with final scores ranging from 0 to 300.

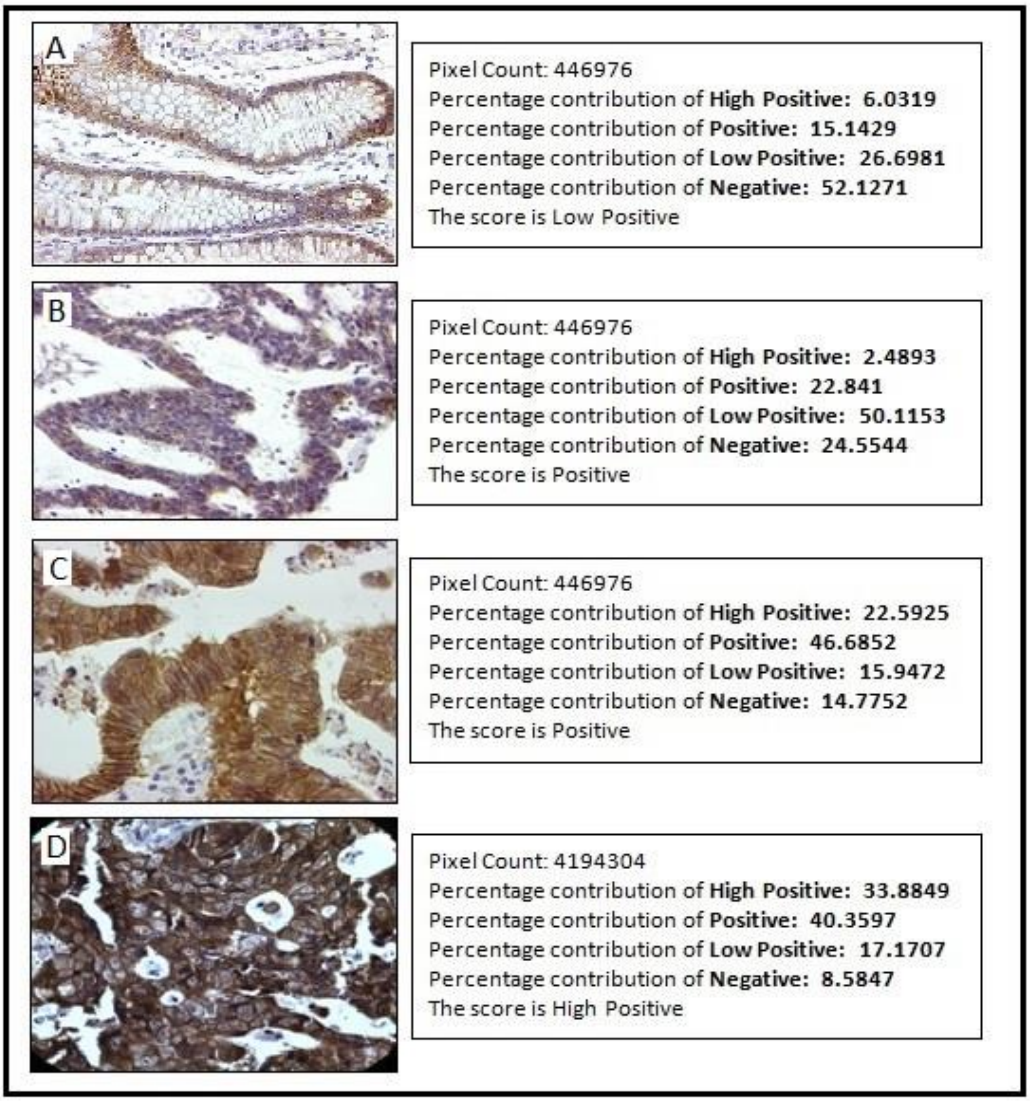


Figure 3.5. Examples of cytoplasmic scoring results produced by IHC Profiler in β-catenin stained tissues (original images are shown). Figure A represents normal colonic epithelium with a low proportion of “positive” and “high-positive” areas, as is expected for normal intestinal mucosa. B, C and D show cancer tissues with different staining patterns. Note that, although B and C exhibit clearly different staining patterns and proportional scores, both are equally classified as “positive” tissues. For this reason, we disregarded the final result and used the proportions to calculate an H-score-like value (continuous variable) using the formula described in the text above.

The last step in the process of testing the suitability of IHC Profiler for the analysis of our biomarkers was to compare the results obtained using this method and our former scoring system for β-catenin (visual modified H-score), as well as other scoring results published in the literature. As will be presented in the next sections, the results obtained with IHC Profiler were concordant with the results from the modified H-score even when comparing samples obtained from different populations (UK and Brazil). These findings encouraged us to use IHC Profiler for the analysis of the cytoplasmic expression of our biomarkers in the Brazilian cohort.

In the original publication, IHC Profiler was used for the evaluation of both cytoplasmic and nuclear markers (Varghese *et al*., 2014). However, in our study, we experienced problems when using the nuclear channel for the evaluation of our proteins. Despite the use of the threshold-setting function in the nuclear channel, the results were discordant compared to the visual inspection of the images and the visual modified H-score. As illustrated in figure 3.6, we found accurate results only when testing markers (not from our candidate list) that were exclusively expressed in the nucleus such as oestrogen receptor. We made contact with the researcher responsible for the development of the tool (Dr. Abhijit De, Molecular Functional Imaging Lab, ACTREC, Tata Memorial Centre, Kharghar, Navi Mumbai, India – personal communication), and he confirmed that the tool could not be precise for "adapting proteins" which move from the cytoplasm to the nucleus and vice versa. Additionally, although there is no consensus in the evaluation of β-catenin nuclear expression, the majority of studies included in a recent meta-analysis of its prognostic significance in CRC measured the percentage of positive nuclei rather than the intensity of staining (Chen *et al*., 2013). The same seems to be the case for most nuclear biomarkers currently used in clinical scenarios such as oestrogen and progesterone receptors (Kurosumi, 2003), and the proliferation marker Ki67 (Mu *et al*., 2015). Therefore, for the nuclear evaluation of our markers, we decided to test another ImageJ plugin: ImmunoRatio.

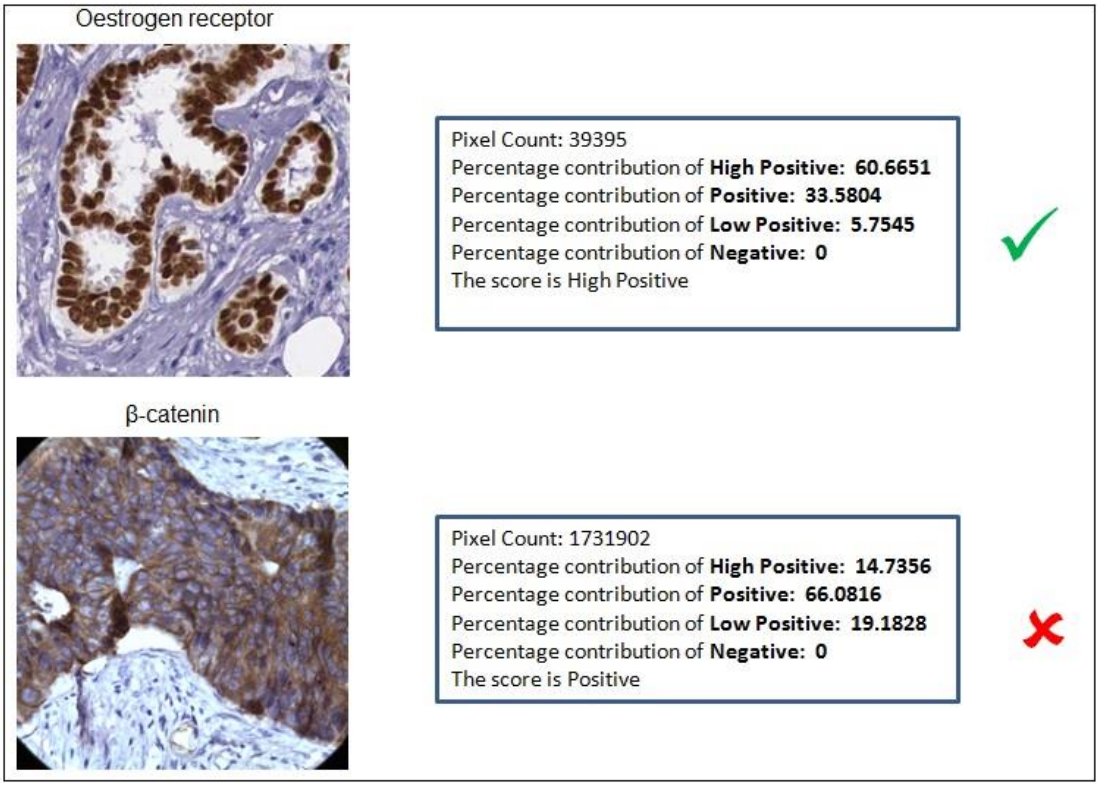


Figure 3.6. IHC Profiler nuclear analysis of oestrogen receptor and β-catenin. The plugin is highly accurate for the assessment of oestrogen receptor expression in breast cancer tissue as it is a nuclear exclusive marker (picture from the Human Protein Atlas, available at www.proteinatlas.org, last accessed in 10/06/17)(Uhlen *et al.*, 2015). For β-catenin, although most nuclei were clearly negative or weakly positive, the result showed high proportions of "high-positive" and "positive" scores (CRC sample). Therefore, IHC Profiler was not considered a suitable tool for the assessment of the nuclear expression of our biomarkers.

Assessing the nuclear expression of biomarkers using ImmunoRatio

ImmunoRatio was developed in 2010 as, according to the authors, "the first publicly available, easily installable web-based application for the immunohistochemical analysis of nuclear markers, such as oestrogen and progesterone receptors and Ki67" (Tuominen *et al*., 2010). These markers were evaluated both visually by pathologists and using the plugin, and a very high correlation coefficient was reported. Later, its use was further tested for Ki67 immunostaining and expanded to other nuclear makers with good accuracy (Sysel *et al*., 2013). The application also uses the colour deconvolution algorithm for DAB-haematoxylin stain separation (Ruifrok and Johnston, 2001, Ruifrok *et al*., 2003) along with several optimising functions

such as blank field correction, background subtraction, adaptive thresholding and nuclear segmentation (Tuominen *et al*., 2010). These functions are illustrated in figure 3.7. Currently, ImmunoRatio can be used as an online tool or it can be freely downloaded as an ImageJ plugin at the University of Tampere (Finland) webpage link http://153.1.200.58:8080/immunoratio/ (last accessed in 26 September 2015).

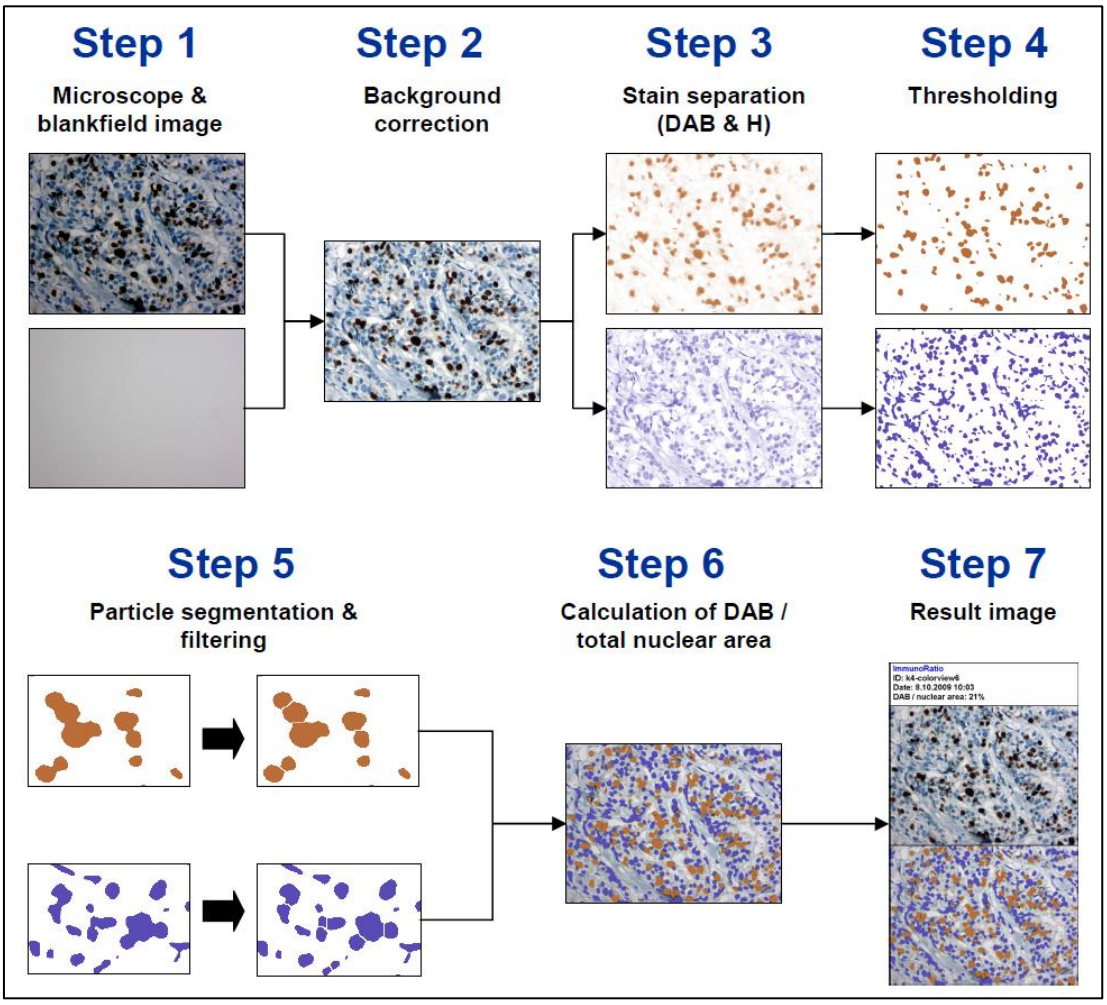


Figure 3.7. ImmunoRatio analysis algorithm. The several steps performed by the application during the assessment of an IHC stained image are summarised here. The final result is presented as a montage showing the original and final images, and the percentage of positive nuclear area. From (Tuominen *et al*., 2010).

Despite the fact that ImmunoRatio was initially tested for nuclear exclusive markers, it can also be used for proteins that are present in both the cytoplasm and nucleus according to the information provided by the

research group that developed this plugin (Jorma Isola MD, PhD; Professor of Cancer Biology, University of Tampere - Finland, BioMediTech – personal communication). Before analysing the images, a threshold optimisation step is necessary. It allows the investigator to fine-tune the sensitivity of the method for the identification of the nuclear areas (based on the haematoxylin staining) and the positively stained areas (DAB staining), as depicted in figure 3.8. This process is usually performed using positive and negative control images, as well as a few test samples in order to assure that the score produced is consistent with the expected visual score for those images. In our study, we carried out the staining of all samples at the same time and using the same conditions for each particular marker. After the “thresholding” was complete, all images were scored using the same settings in order to guarantee comparability among the samples. As different makers were assessed following separate staining procedures, the process of thresholding was repeated accordingly for each protein.

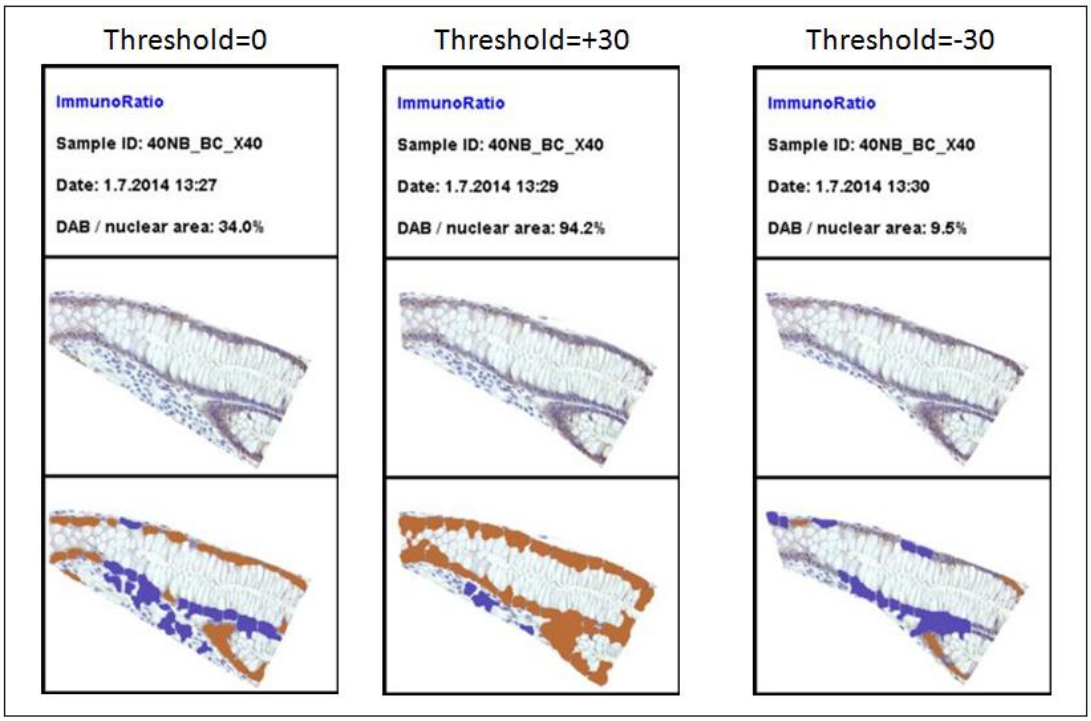


Figure 3.8. The sensitivity of the plugin can be fine-tuned using the built-in thresholding capability. These figures illustrate different results obtained when different DAB threshold values were applied. The same can be done for haematoxylin staining (not shown).

After the optimisation step, the samples from the Brazilian cohort of patients that had been stained for β-catenin were scored using ImmunoRatio and the results were compared with the modified H-score obtained from the UK-cohort and with the published literature. ImmunoRatio results showed high nuclear localisation of β-catenin in neoplastic tissues compared with the normal mucosa, thus reflecting exactly what was found using the modified H-score and what is expected based on the literature (results will be presented in the next section). Therefore, we also used ImmunoRatio to score the other candidate biomarkers in the Brazilian cohort.

3.3. Evaluation of β-catenin immunostaining confirms Wnt pathway activation in neoplastic tissues

In this section, the results obtained during the analysis of β-catenin stained samples will be described. Data will be presented according to the hypotheses that were formulated. As explained in the previous sections, the results from the evaluation of the UK-cohort were obtained using the manual (or visual) modified H-score, whereas the results for the Brazilian cohort were obtained by electronic scoring using IHC Profiler and ImmunoRatio. When available, the results for both cohorts are compared. As the manual scoring results in a single value, we present it using a box-plot chart. However, the electronic scoring produced two values per image (cytoplasmic and nuclear scores separately). Therefore, we decided to use a dual bar-and-line graph with different colours to allow simultaneous visualisation of the results. Figure 3.9 provides an example of β-catenin staining pattern in a CRC sample and the adjacent intestinal epithelium.

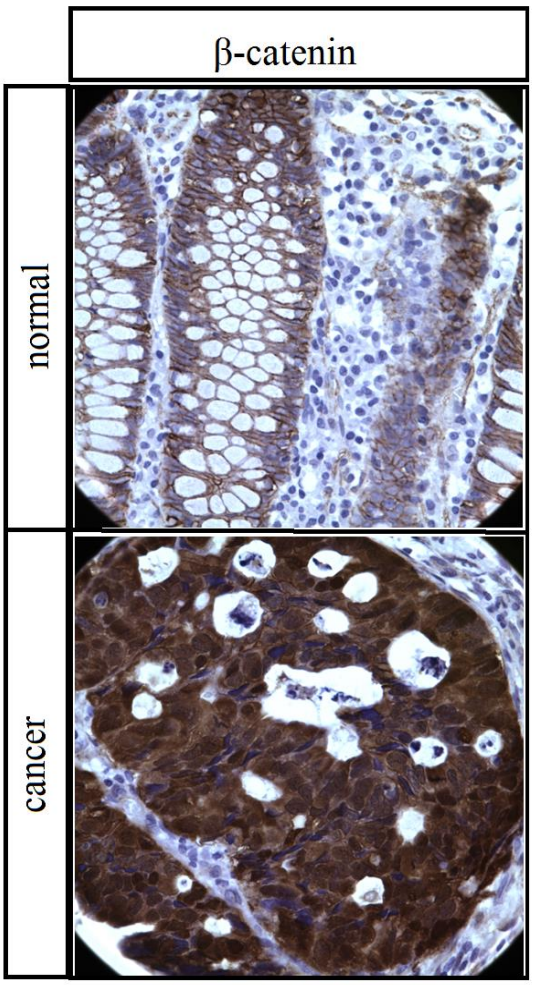


Figure 3.9. An example of β-catenin staining pattern. A predominantly membranous localisation is observed in the normal epithelium, whereas a strong cytoplasmic and nuclear staining is seen in the cancer cells. These tissues were collected from a patient with stage IV (Dukes' stage D) CRC. Magnification: 630x.

Assessment of β-catenin expression in adenomas, cancer and adjacent tissues – UK-cohort

For the evaluation of the UK-cohort, we divided the samples into 4 groups: adjacent non-neoplastic colonic mucosa, low-grade adenomas, high-grade adenomas and invasive cancers. The aim of this analysis was to confirm Wnt pathway activation in our samples, therefore validating our scoring tools for the analysis of the remaining biomarkers. As can be noted when comparing these groups with those used in the preliminary testing and optimisation steps of the manual and electronic scoring methods (see figure 3.2 in the previous section), we decided to make some modifications. The "polyp cancer" group - invasive cancers incidentally found after a polypectomy, was excluded due to the low number of samples and the heterogeneity of these cases in terms of staging. Additionally, the various

cancer groups were amalgamated into a single group, as there were no apparent differences in the expression of the biomarkers across different cancer stages. As depicted in figure 3.10, the results showed a clear increase in modified H-score from the adjacent non-neoplastic colonic mucosa to the low-grade adenoma group, and this increase was sustained in more advanced neoplastic lesions (high-grade adenomas and invasive cancers). This denotes a translocation of the protein from the cell membrane into the cytoplasm and the nucleus and is a surrogate marker of Wnt pathway activation.

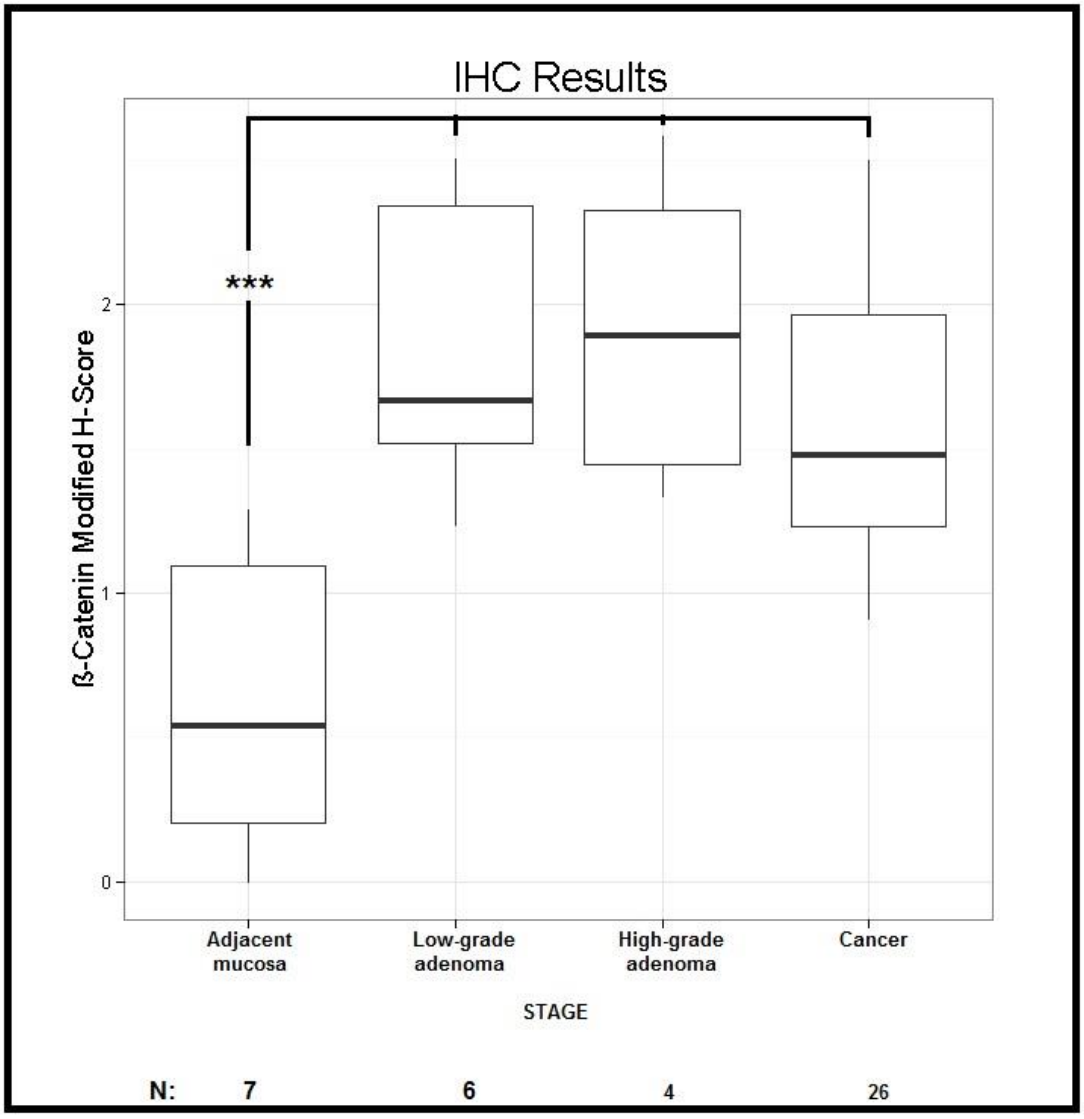


Figure 3.10. Modified H-scores for β-catenin in the UK-cohort. A clear increase in cytoplasmic and nuclear protein localisation is seen from the earliest neoplastic stages, which is maintained in more advanced and invasive lesions. *** $p<0.001$ (Kruskal-Wallis test followed by Mann-Whitney U test for post-hoc pair-wise comparisons and Bonferroni correction).

Assessment of β-catenin expression in cancer versus adjacent tissues – Brazilian cohort

For the evaluation of the Brazilian cohort, we initially had only adjacent normal mucosa and tumour tissues. As the number of samples was higher compared to the UK-cohort, we decided to test again whether or not there was any difference in β-catenin staining across the different cancer stages. For this purpose, we split the samples into 2 groups: early stage, encompassing stages I and II (corresponding to Dukes' stages A and B) and late stage, including stages III and IV (Dukes' stages C and D).

Despite the use of different scoring methods (as explained earlier) and a different patient population, the results exhibited in figure 3.11 showed the same pattern as was observed in the UK-cohort. A clear and statistically significant increase in both nuclear and cytoplasmic localisation of β-catenin was observed in cancer tissues compared to the adjacent mucosa. No difference was however seen when comparing early-stage versus late-stage cancer groups. These findings provided strong support for the use of this electronic scoring system to analyse the expression of the other biomarkers that we subsequently tested.

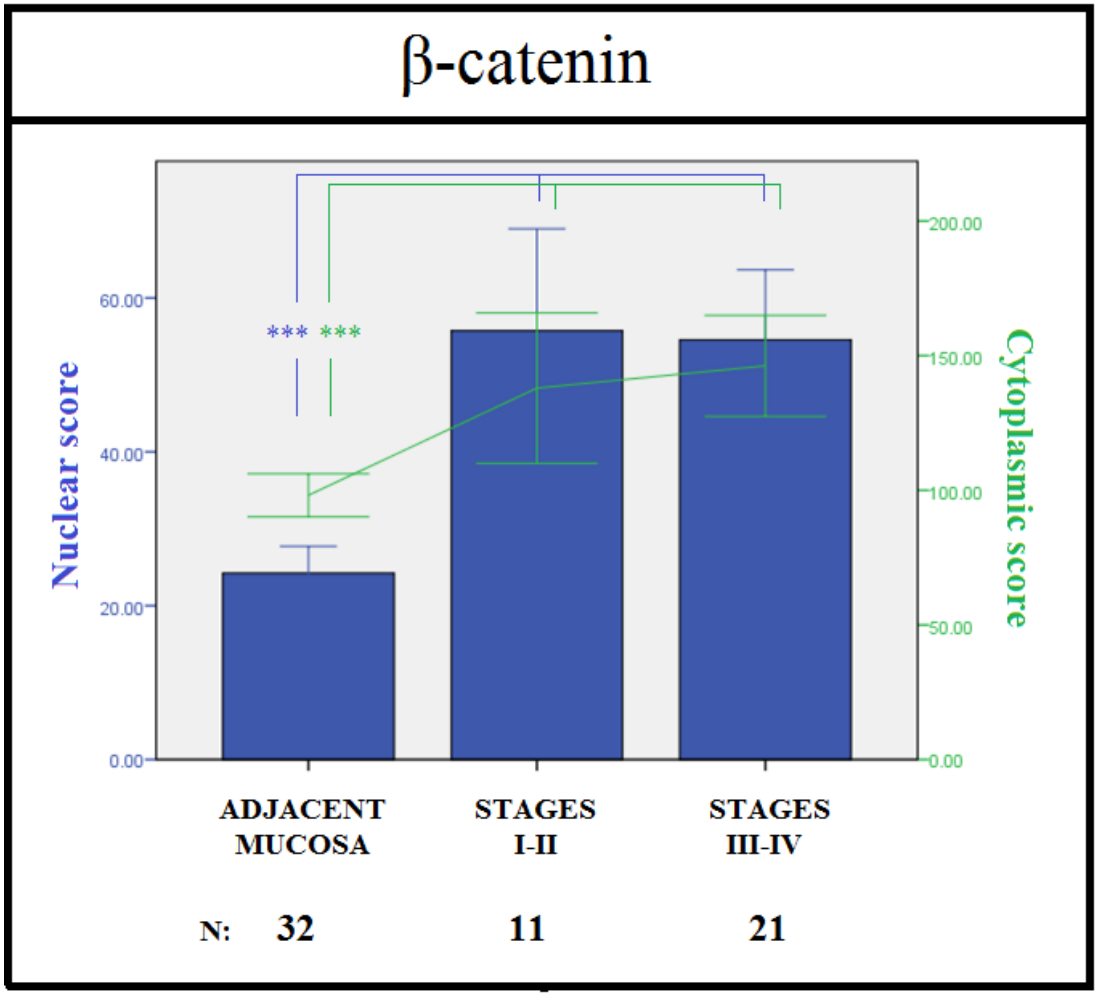


Figure 3.11. Electronic scoring of the Brazilian cohort stained for β-catenin. Blue bars represent the mean nuclear score obtained using the ImmunoRatio plugin. Green line represents the mean cytoplasmic score produced by the IHC Profiler plugin. Additional blue and green lines and asterisks show statistically significant differences between groups. Marked increases in both nuclear and cytoplasmic scores were observed in cancer groups compared to the adjacent mucosa. No difference was seen between different cancer stages. *** $p<0.001$ (Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons). Error bars represent ± 2 SE. N = sample numbers.

Comparison between normal mucosa from control individuals and adenoma tissues

The results from the UK-cohort regarding β-catenin immunostaining suggested that protein translocation from membrane into the inner parts of the cells occurs early during the adenoma-carcinoma sequence. The initial set of Brazilian samples encompassed cancer and adjacent normal samples

only. Therefore, in order to validate the results obtained initially, we decided to expand the collection of samples to include colorectal adenomas. Additionally, instead of comparing the adenomas with adjacent tissues, we also collected mucosa samples from individuals who had no colonoscopic evidence of intestinal lesion (normal controls). Most experimental studies using human samples compare neoplastic tissues with the apparently normal adjacent mucosa only. However, in many cases, cancer is the result of the effect of carcinogenic factors that affect large areas of the organ, a process known as “field cancerisation”. This is particularly true for epithelial cancers which commonly arise as a result of local factors. Some reports have shown that the transcriptomic profile of tissues adjacent to different tumour types is different from the normal epithelium from individuals without cancer, including CRC (Sanz-Pamplona et al., 2014, Chandran et al., 2005, Raudenska et al., 2015b). Therefore, assessing a cohort of normal control samples might reinforce the conclusions of our study.

In this analysis, we included 10 samples from normal control patients and 18 samples of colorectal adenomas. Only 2 of the 18 adenomas were high-grade lesions. The remaining 16 cases exhibited low-grade dysplasia. Hence, we combined all the adenomas into a single group for the purposes of comparison. Whole-sections were stained for β-catenin as described in *Methods* and the analysis and scoring were carried out using the electronic plugins described above (IHC Profiler and ImmunoRatio). As depicted in figure 3.12, the assessment of β-catenin immunostaining in these samples showed that nuclear localisation of the protein was increased in adenomas when compared to normal epithelium. The cytoplasmic staining seemed to be increased as well, although this was not statistically significant. These findings reflect what was observed with the UK-cohort and reinforce the concept that Wnt pathway activation is an early event during colorectal carcinogenesis.

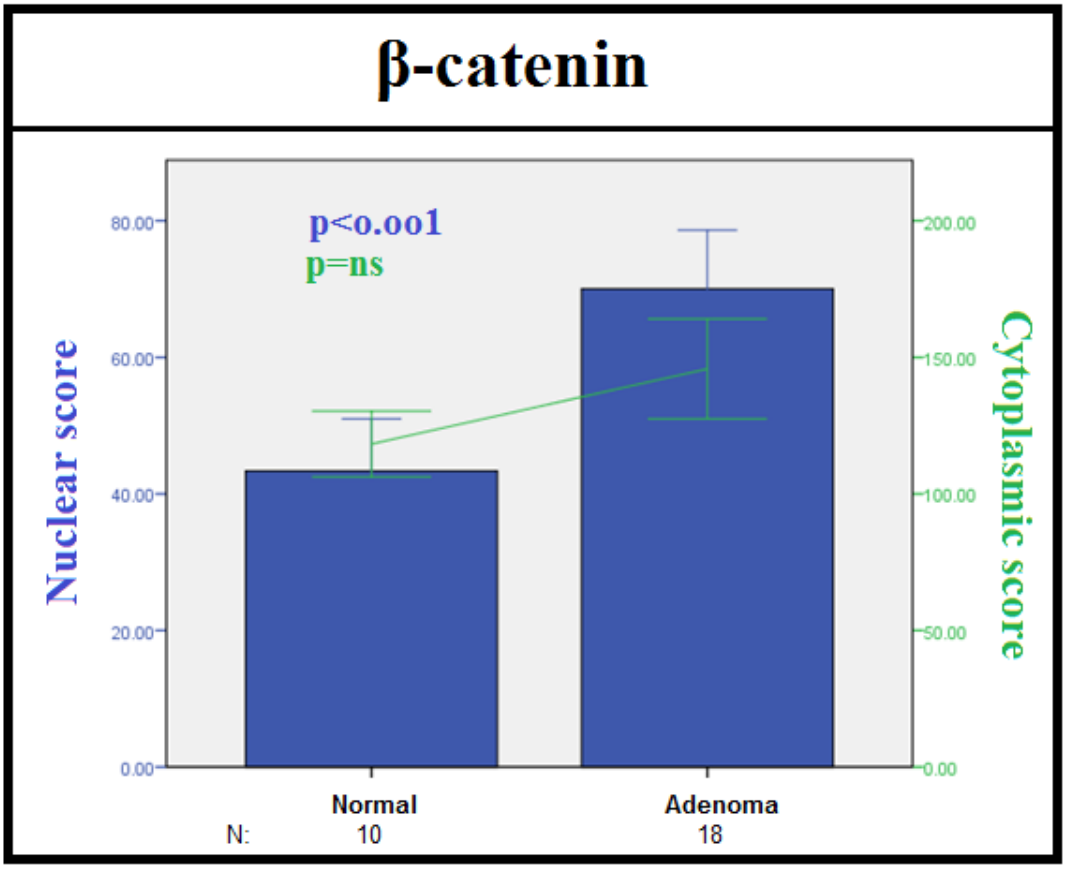


Figure 3.12. Analysis of β-catenin immunostaining in normal and adenoma samples from the Brazilian cohort of patients. The nuclear score was significantly higher in the adenoma group, denoting Wnt pathway activation at this early time point in the adenoma-carcinoma sequence (p values yielded by Mann-Whitney U test). There was also a trend towards an increase in cytoplasmic staining in the adenoma group, but this was not statistically significant. Error bars represent ± 2 SE. N = sample numbers.

The preliminary analysis of β-catenin expression in these tissues permitted us to confirm that Wnt signalling pathway is activated in all neoplastic groups – a finding consistent with several previous reports (Yoshida *et al*., 2015, Abdelmaksoud-Damak *et al*., 2015, Kobayashi *et al*., 2000) and with a large published meta-analysis (Chen *et al*., 2013). At the same time, the results support the use of the scoring systems that we have subsequently employed for the other candidate biomarkers. Below, the results for the remaining candidate proteins are presented. Both cohorts (UK and Brazil) were used for the evaluation of NAP1L1 and RPL6. The analysis of PHB, HMGB1, SFRS2 and CDC5L was limited to the Brazilian cohort due

to the shortage of target tissues in the TMA block produced from the UK-cohort, as explained previously.

3.4. Expression of NAP1L1 in the adenoma-carcinoma sequence

Sections from the UK-cohort TMA were stained for NAP1L1 using the protocol and conditions described in *Chapter 2*. Scanned slides were analysed according to the procedures explained earlier, during the description of the manual modified H-score. Figure 3.13 shows an example of the NAP1L1 staining pattern in normal and cancer tissues. The results from this analysis are shown in figure 3.14. A decrease in the median modified H-score was observed in the high-grade adenoma group; this was also observed in the invasive cancer group. Interestingly, and different from what was observed for β-catenin, we did not notice any difference in NAP1L1 immunostaining between the adjacent mucosa and the low-grade adenoma group. This suggests that the displacement of the protein from the nuclei of the cells occurs only when the dysplastic process progresses into a more aggressive phenotype.

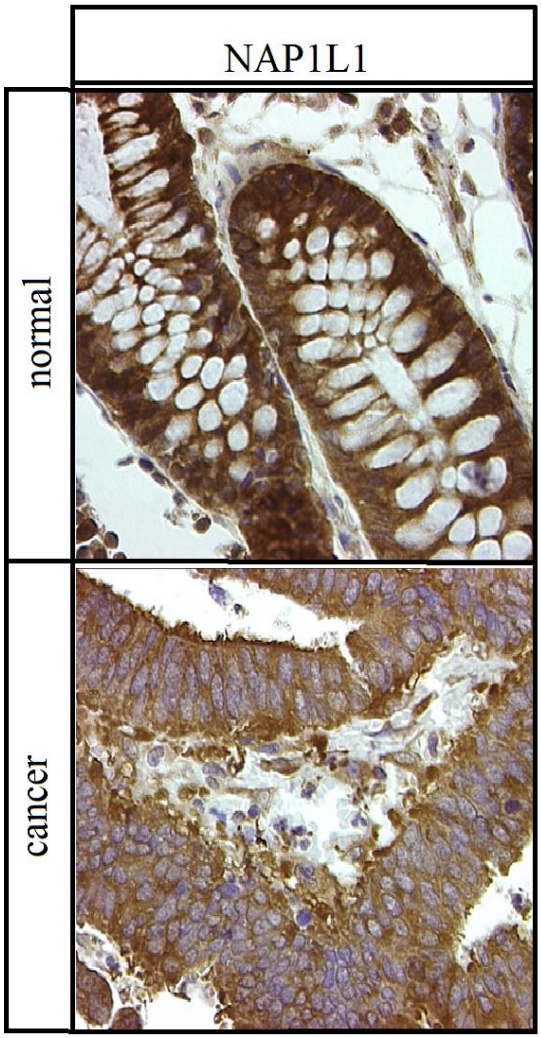


Figure 3.13. An example of NAP1L1 staining pattern. Normal epithelium shows strong nuclear and cytoplasmic staining. Cancer tissues exhibit a decrease in the nuclear content of the protein. The cytoplasmic staining is also decreased, although to a minor degree in this case. These tissues were collected from a patient with stage III (Dukes' stage C) CRC. Magnification: 630x.

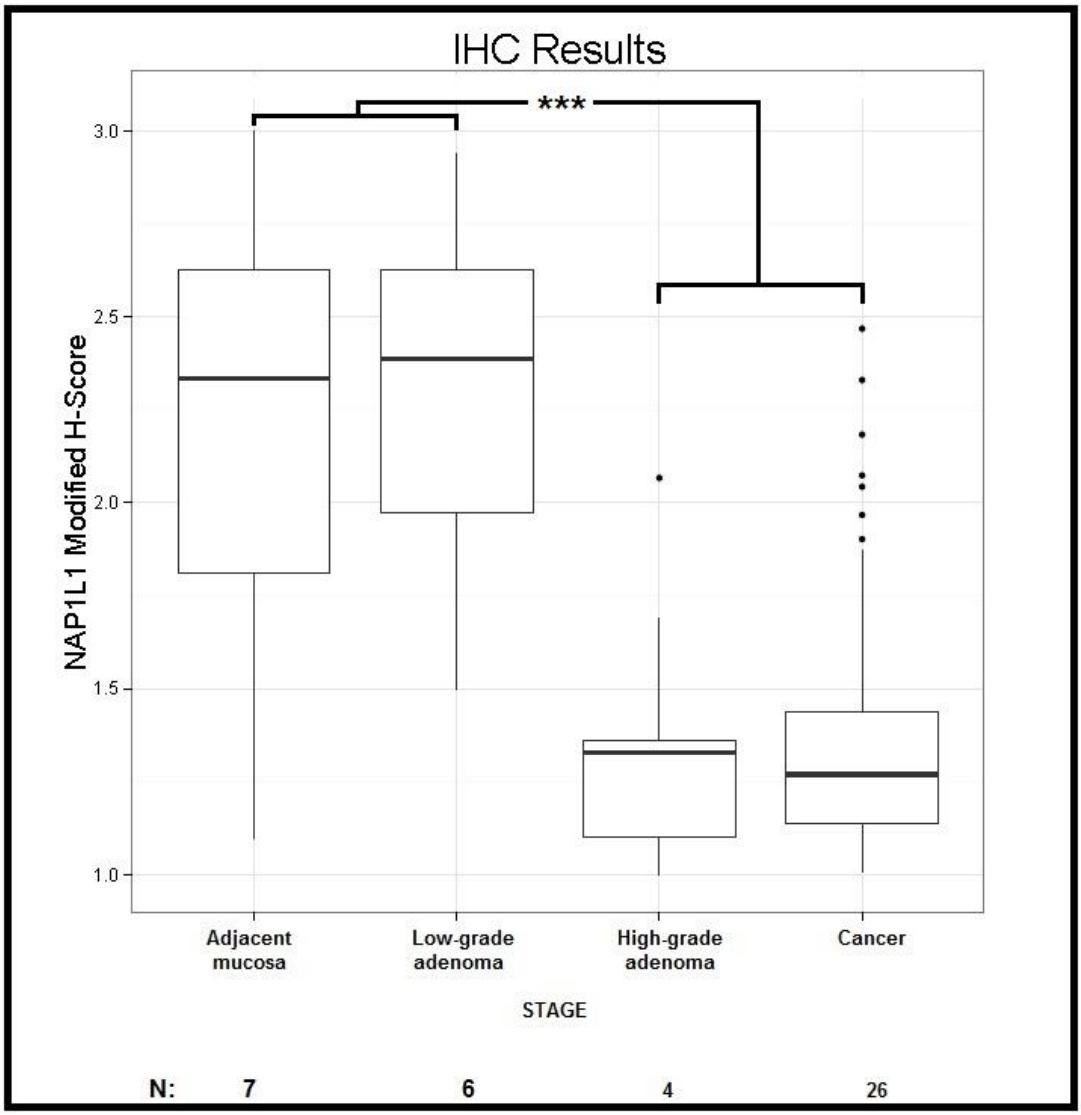


Figure 3.14. Modified H-scores for NAP1L1 in the UK-cohort. A clear decrease in cytoplasmic/nuclear protein localisation is seen in the high-grade adenoma group and in the invasive cancer group, compared to both the adjacent mucosa and the low-grade adenoma groups. *** $p<0.01$ for adjacent vs high-grade; $p<0.001$ for adjacent vs cancer. $p<0.001$ for low-grade vs high-grade and vs cancer (Kruskal-Wallis test followed by Mann-Whitney U test for post-hoc pair-wise comparisons and Bonferroni correction). N = sample numbers.

To further investigate these results, whole-sections from blocks included in the Brazilian cohort were also stained for NAP1L1 and were scored using the electronic plugins. The results using this alternative cohort confirmed the findings from the UK-cohort, and showed a decrease in both the nuclear and the cytoplasmic expression of the protein in cancer samples (see figure 3.15). Again, no difference was seen between different cancer stage groups. Taken together, these findings provide strong evidence for an

important role of NAP1L1 content and localisation in colorectal carcinogenesis.

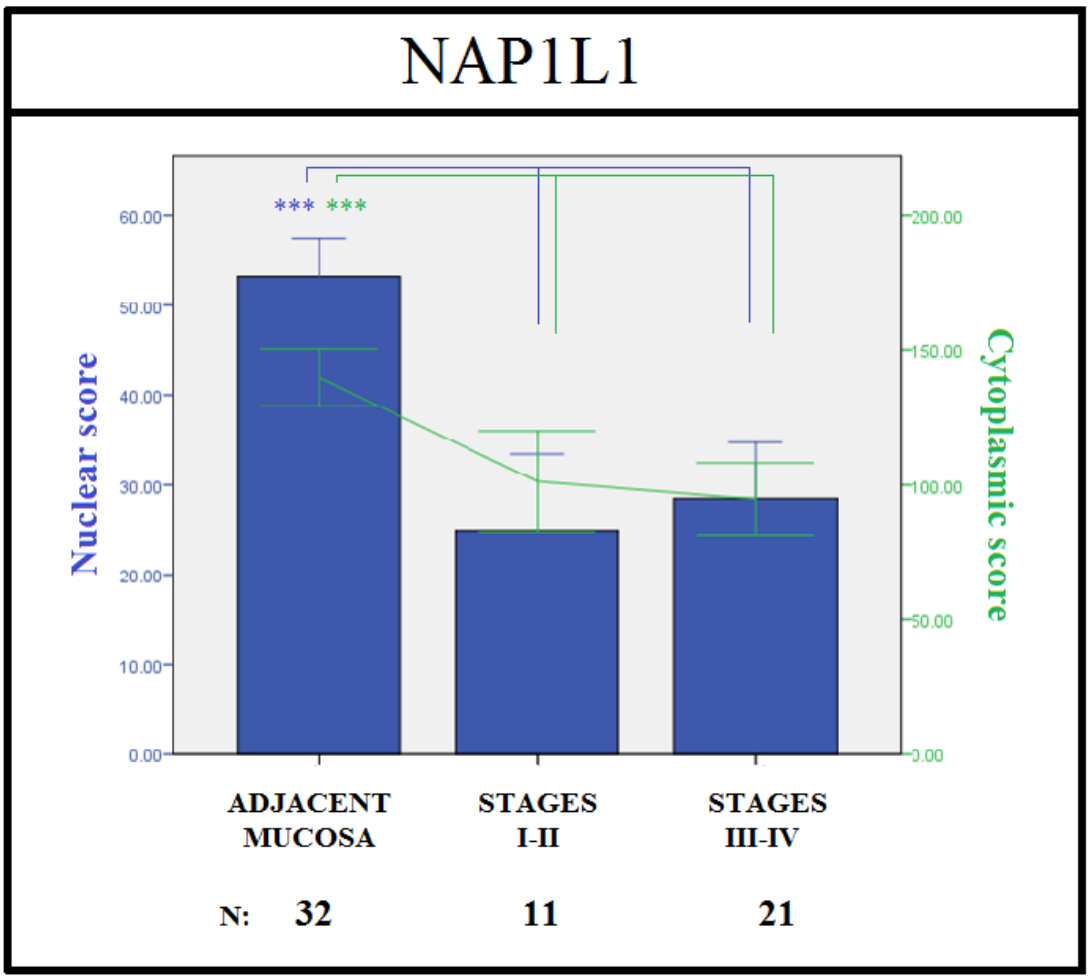


Figure 3.15. Electronic scoring of the Brazilian cohort stained for NAP1L1. Both the nuclear and the cytoplasmic scores were decreased in the cancer groups compared to the adjacent mucosa. No difference was observed between different stages of cancer. *** p<0.001 (Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons). Error bars represent ± 2 SE. N = sample numbers.

To assess how early this alteration occurred and to confirm the results from the UK-cohort, we also assessed the NAP1L1 immunostaining pattern in normal mucosa samples from control individuals and in the adenoma samples included in the Brazilian cohort. As depicted in figure 3.16, no difference was observed in protein content or localisation between normal tissues and adenomas. This result mirrors the patterns observed in adjacent

tissues and low-grade adenomas in the UK-cohort. Unfortunately, the low number of high-grade adenomas in the Brazilian cohort did not allow us to compare this group with the others. However, if the results from the UK-cohort are correct, these findings might support a concept of using NAP1L1 immuno-localisation as a biomarker for the conversion of low-grade into high-grade dysplasia.

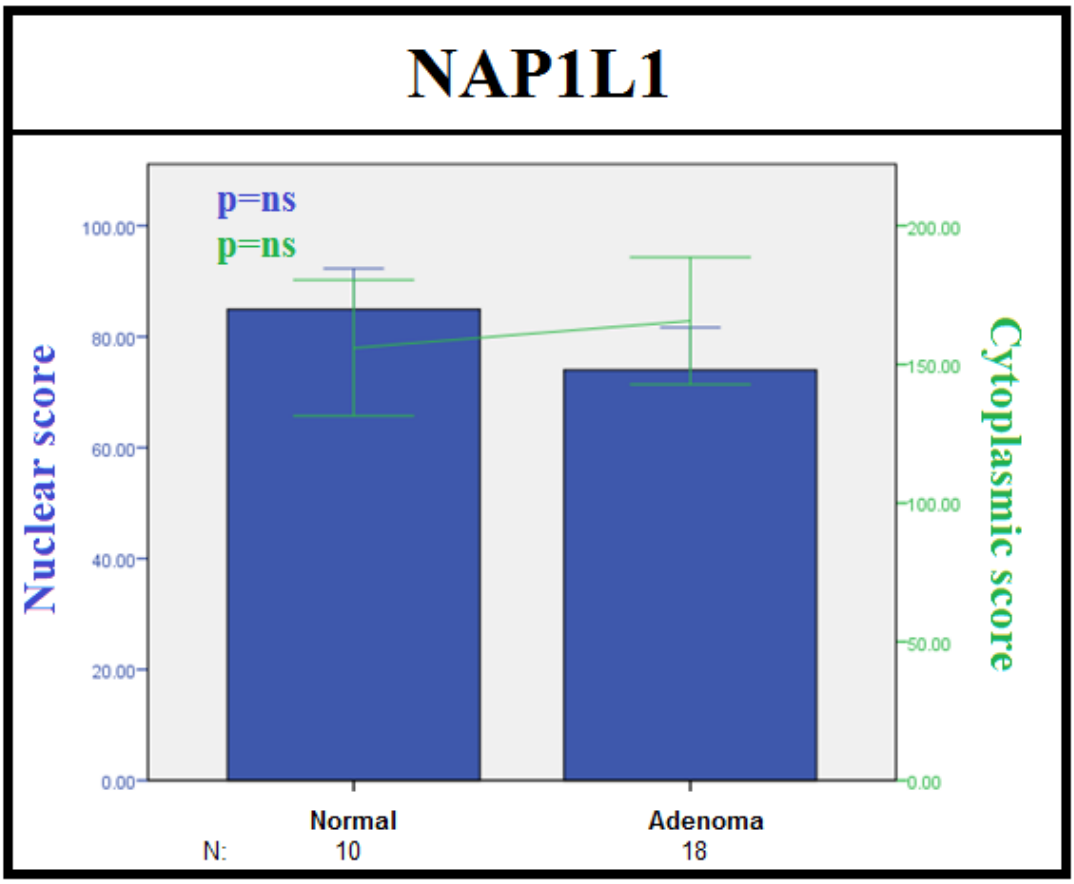


Figure 3.16. Analysis of NAP1L1 immunostaining in normal and adenoma samples from the Brazilian cohort. No difference was observed between normal and adenoma samples (Mann-Whitney U test). Error bars represent ± 2 SE. N = sample numbers.

Most of the scientific literature regarding NAP1L1 expression in cancer tissues is limited to intestinal and pancreatic neuroendocrine tumours, a group of neoplasms which have different genotypic and phenotypic profiles compared to CRC (Schimmack *et al*., 2014, Kidd *et al*., 2006). The only report of the analysis of NAP1L1 in human CRC that we have found was a

study published in 2002 showing that *NAP1L1* mRNA expression was increased in tumours relative to the adjacent mucosa (Line *et al*., 2002). No immunohistochemical study evaluating this protein in CRC tissues has been found. Therefore, our results provide the first description of the NAP1L1 immunostaining pattern in CRC and highlight the potential importance of this protein during colorectal carcinogenesis.

3.5. Expression of RPL6 in the adenoma-carcinoma sequence

Following the same procedures used for the assessment of the previous biomarkers, we analysed the immunohistochemical expression of RPL6 in samples from both the UK and Brazilian cohorts. The staining pattern (figure 3.17) and results (figure 3.18) for the UK-cohort were highly consistent with the pattern observed for NAP1L1. A clear-cut difference in RPL6 immuno-localisation was observed between adjacent/low-grade dysplastic adenoma groups versus high-grade dysplastic adenoma/cancer groups. This suggests that this protein could also possibly differentiate between low-risk and high-risk polyps (high-grade dysplasias), besides being a cancer maker.

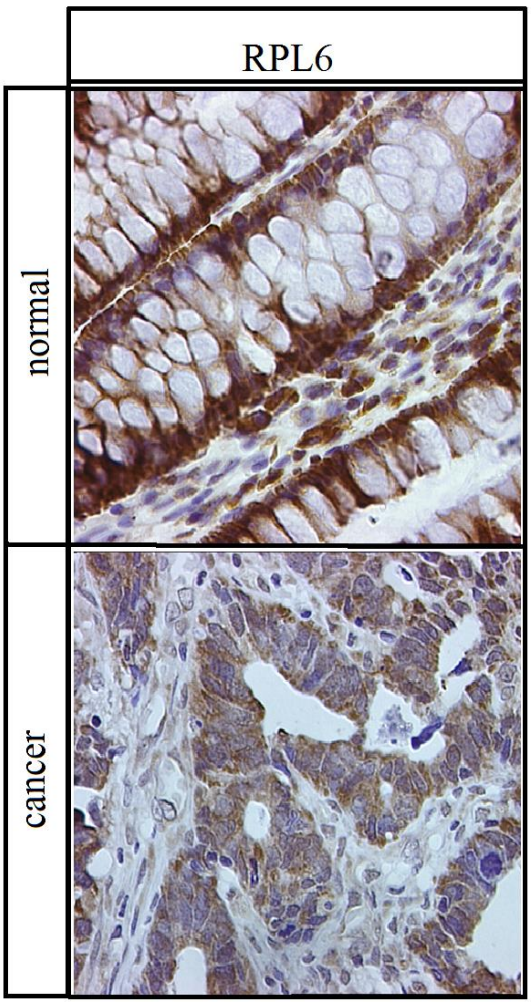


Figure 3.17. An example of RPL6 staining pattern. Cancer tissues exhibited decreased levels of staining in both the cytoplasm and the nucleus of the cells, a similar pattern to NAP1L1 immuno-staining. These tissues were collected from a patient with stage I (Dukes' stage A) CRC. Magnification: 630x.

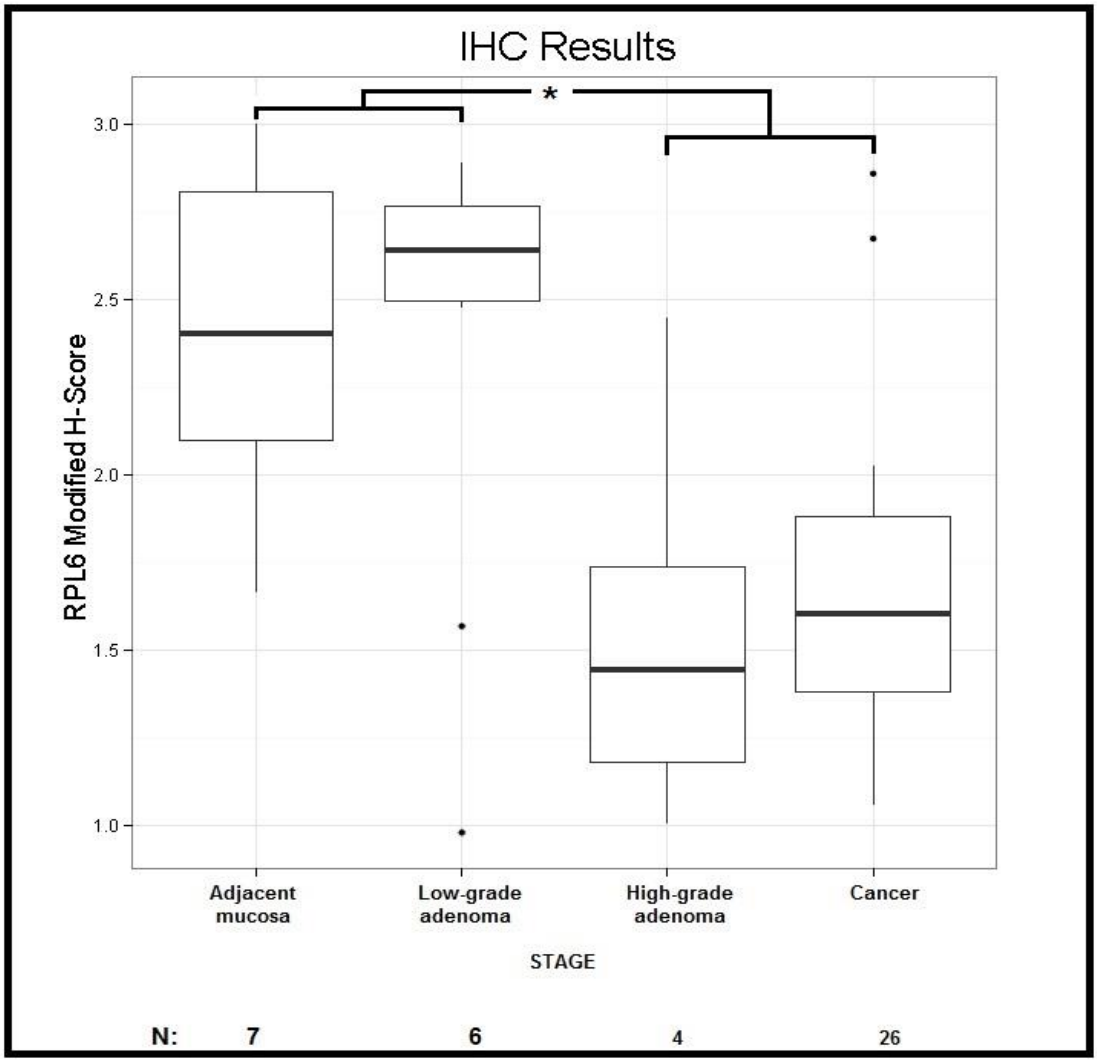


Figure 3.18. Modified H-scores for RPL6 in the UK-cohort. As was observed with NAP1L1, a decrease in cytoplasmic/nuclear protein localisation was seen in the high-grade adenoma group and in the invasive cancer group, compared to both the adjacent mucosa and the low-grade adenoma groups. * p=0.05 for adjacent vs high-grade; p<0.001 for adjacent vs cancer; p=0.05 for low-grade vs high-grade; p=0.01 for low-grade vs cancer (Kruskal-Wallis test followed by Mann-Whitney U test for post-hoc pair-wise comparisons and Bonferroni correction). N = sample numbers.

As before, we also tested the Brazilian cancer cohort using IHC Profiler and ImmunoRatio for cytoplasmic and nuclear scoring, respectively. As can be seen in figure 3.19, the nuclear analysis showed a statistically significant decrease in RPL6 immuno-expression in cancer tissues compared to the adjacent non-neoplastic colon group. However, the magnitude of the difference was smaller when compared to the results obtained with the manual modified H-score used in the UK-cohort. Although it is theoretically

possible that different populations exhibit different patterns of biomarker expression (as explained in *Chapter 1 – Introduction*), our hypothesis is that the electronic tool is not sensitive enough to detect small differences in staining for this marker. The cytoplasmic scores were similar between all the groups tested. This finding does not necessarily disagree with the results obtained for the UK-cohort, as the modified H-score used in that analysis is mainly influenced by the nuclear rather than the cytoplasmic score.

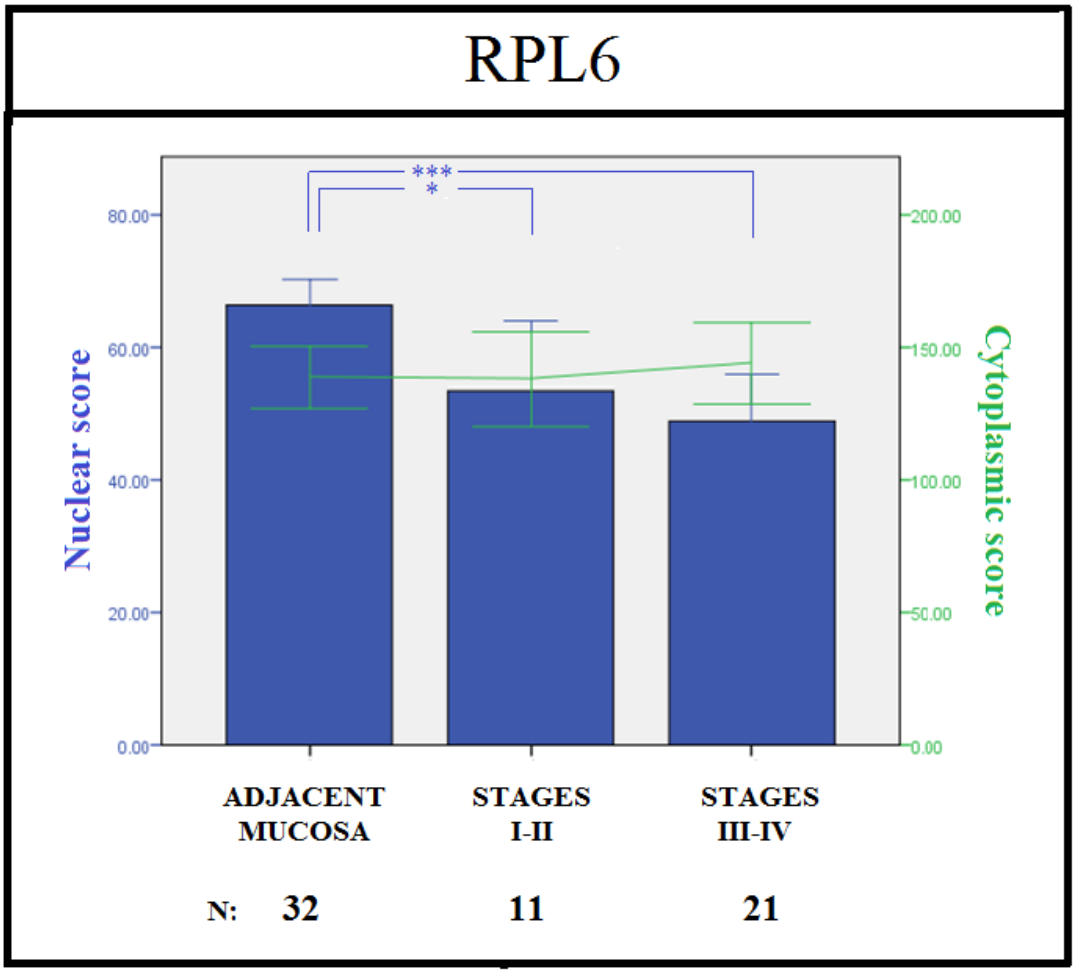


Figure 3.19. Electronic scoring of the Brazilian cohort stained for RPL6. A decrease in nuclear scores was observed in cancer tissues compared to the adjacent mucosa group. Different stage groups exhibited similar scores. No difference was observed in the cytoplasmic immunostaining score between the groups. ***p<0.001, *p=0.05 (Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons). Error bars represent ± 2 SE. N = sample numbers.

Next, we performed the evaluation of RPL6 immuno-expression in normal samples and adenomas from the Brazilian cohort of patients. Again, the results were concordant with the analysis of NAP1L1. Neither the cytoplasmic nor the nuclear staining scores showed any difference between normal mucosa and adenomas. The results are presented in figure 3.20.

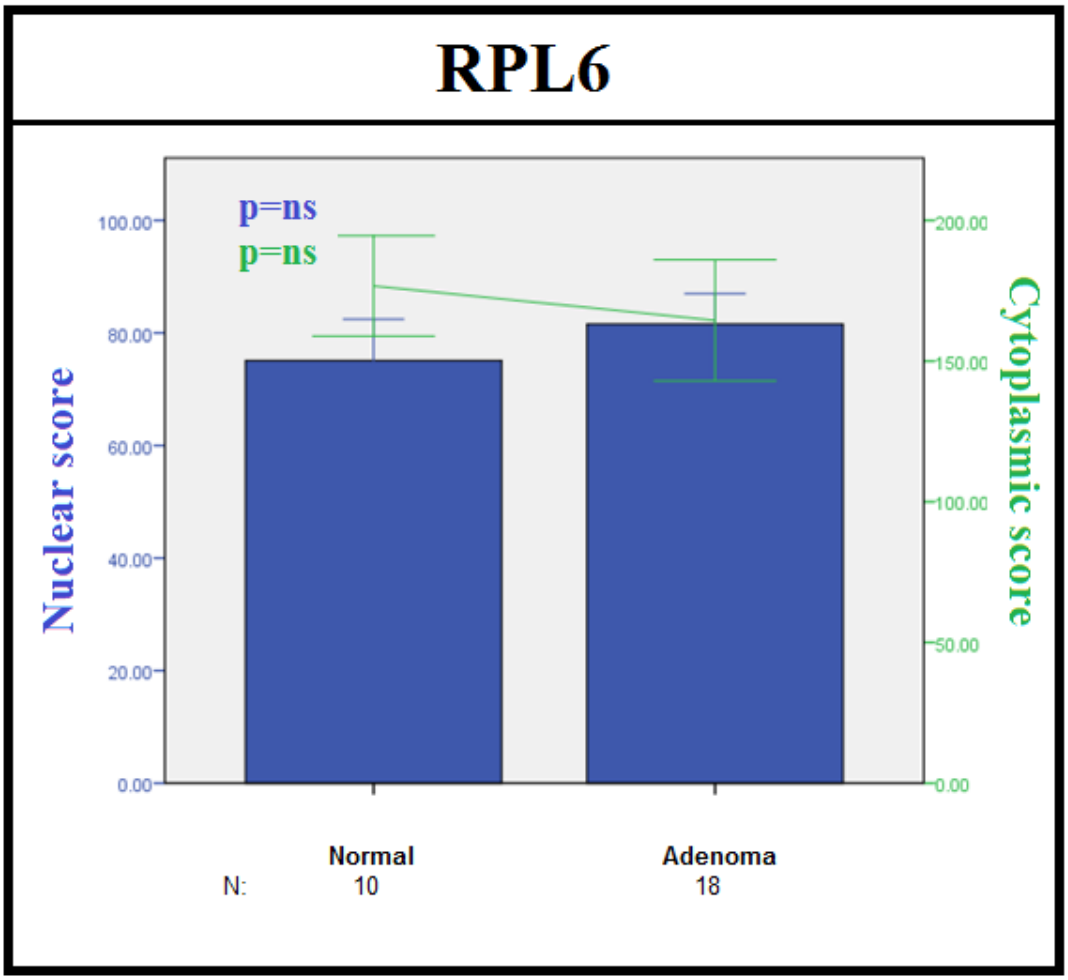


Figure 3.20. Analysis of RPL6 immunostaining in normal and adenoma samples from the Brazilian cohort. As observed with NAP1L1, no difference was found between normal and adenoma samples (Mann-Whitney U test). Error bars represent ± 2 SE. N = sample numbers.

The only type of cancer in which RPL6 expression has been well studied is gastric adenocarcinoma (Wu *et al*., 2011, Gou *et al*., 2010, Du *et al*., 2005). An elevated gene expression was initially demonstrated in multidrug-resistant gastric cancer cell lines (Du *et al*., 2005). Later, the same

group showed that RPL6 immuno-expression was negative in normal gastric mucosa, positive in the cytoplasm of the majority of gastric cancer tissues and only rarely present in the nuclei of cancer cells (Gou *et al*., 2010). In our study of the expression of RPL6 in CRC, we found a different result, with normal colorectal epithelium exhibiting strongly positive immunostaining (especially in the nucleus). However, in CRC cancer tissues, RPL6 immuno-expression was apparently similar to that previously described in gastric cancer.

Taken together, these results suggest that both NAP1L1 and RPL6 have similar cellular distributions in normal, pre-malignant and malignant colorectal tissues. Both proteins exhibited a differential staining pattern from the high-grade dysplastic adenoma stage (as opposed to β-catenin which showed nuclear translocation from the low-grade adenoma stage) suggesting that these markers may play a role in the acquisition of a more aggressive phenotype by the adenoma. These findings definitely warrant further studies about the role of these proteins during the colorectal carcinogenesis.

3.6. Expression of Prohibitin (PHB) in CRC tissues

The remaining candidates were not assessed in the UK-cohort for the reasons previously explained (insufficient samples). Therefore, an analysis of the immuno-expression of PHB in the Brazilian cohort is now described. Staining procedures were performed according to the protocol and conditions specified in *Chapter 2*. Figure 3.21 shows an example of PHB immunostaining in normal adjacent and tumourous colonic tissues. This protein also exhibited a decrease in nuclear expression in cancer tissues compared with the adjacent unaffected mucosa.

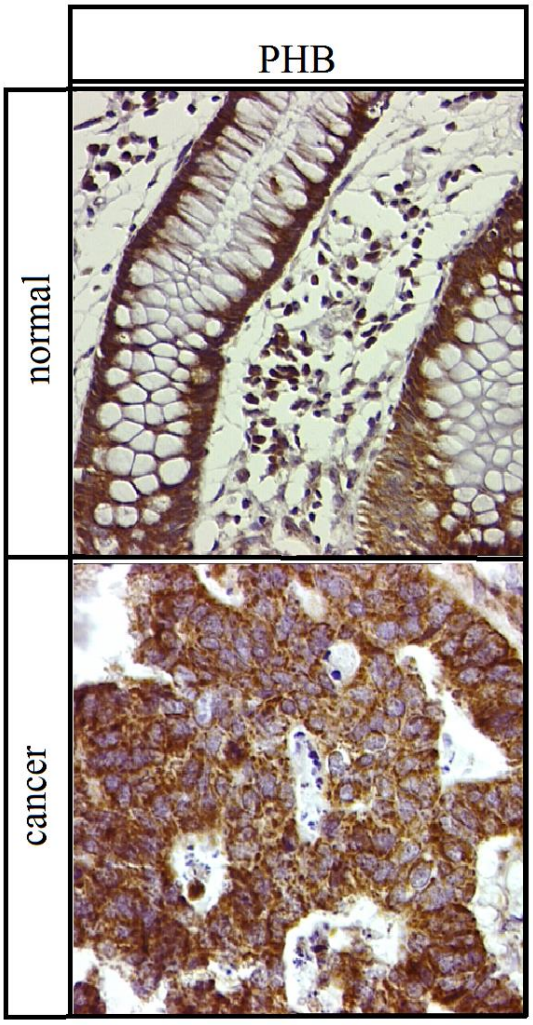


Figure 3.21. An example of PHB staining pattern. Normal tissues generally exhibited positive nuclear and cytoplasmic staining, whereas cancer cells showed decreased nuclear staining whilst keeping cytoplasmic positivity. These tissues were collected from a patient with stage III (Dukes' stage C) CRC. Magnification: 630x.

The electronic evaluation carried out with the plugin ImmunoRatio confirmed this observation (figure 3.22). A statistically significant decrease in the nuclear score was observed in cancer tissues compared with the adjacent mucosa. However, results obtained with IHC Profiler showed that the cytoplasmic immuno-expression of PHB was actually increased in malignant cells (figure 3.22). This increase was not easily perceived by visual inspection of the images. If correct, these findings suggest that the protein is displaced from the nucleus into the cytoplasm during malignant transformation.

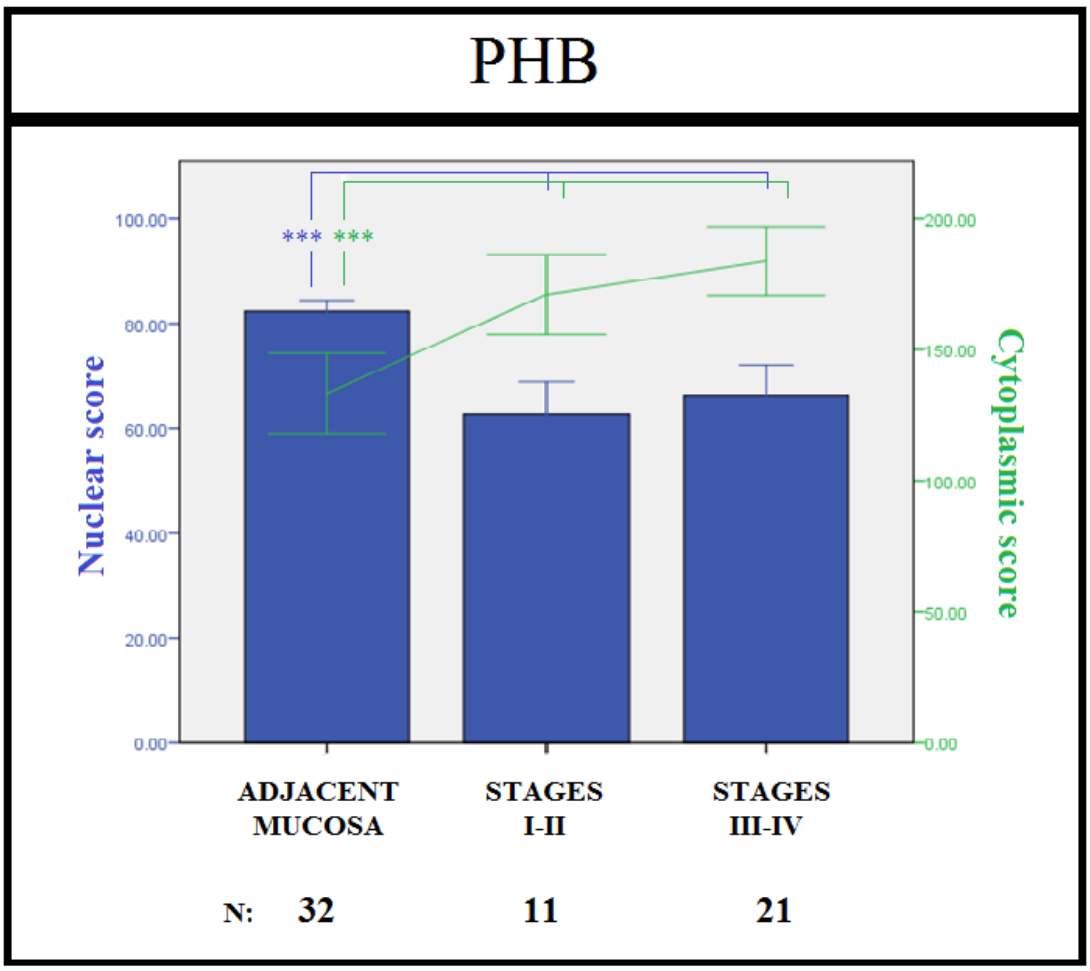


Figure 3.22. Electronic scoring of the Brazilian cohort stained for PHB. A decrease in nuclear scores was observed in cancer tissues compared to the adjacent mucosa group whereas an increase in the cytoplasmic score was seen. Different cancer stage groups exhibited similar scores. ***p<0.001 (Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons). Error bars represent ± 2 SE. N = sample numbers.

To assess how early these alterations occur, we also analysed normal mucosa from healthy controls and adenoma samples from polyp-bearing individuals. As illustrated in figure 3.23, although the nuclear score exhibited a trend towards decreased expression in adenomas, this did not reach statistical significance. No difference was observed in the cytoplasmic score. As was the case for NAP1L1 and RPL6, this suggests that the expression of this protein is not altered in benign low-grade adenomas and it specifically identifies malignant tissues. We could not assess whether or not this protein was altered in high-grade adenomas due to the low number of such cases in the Brazilian cohort (2 cases only).

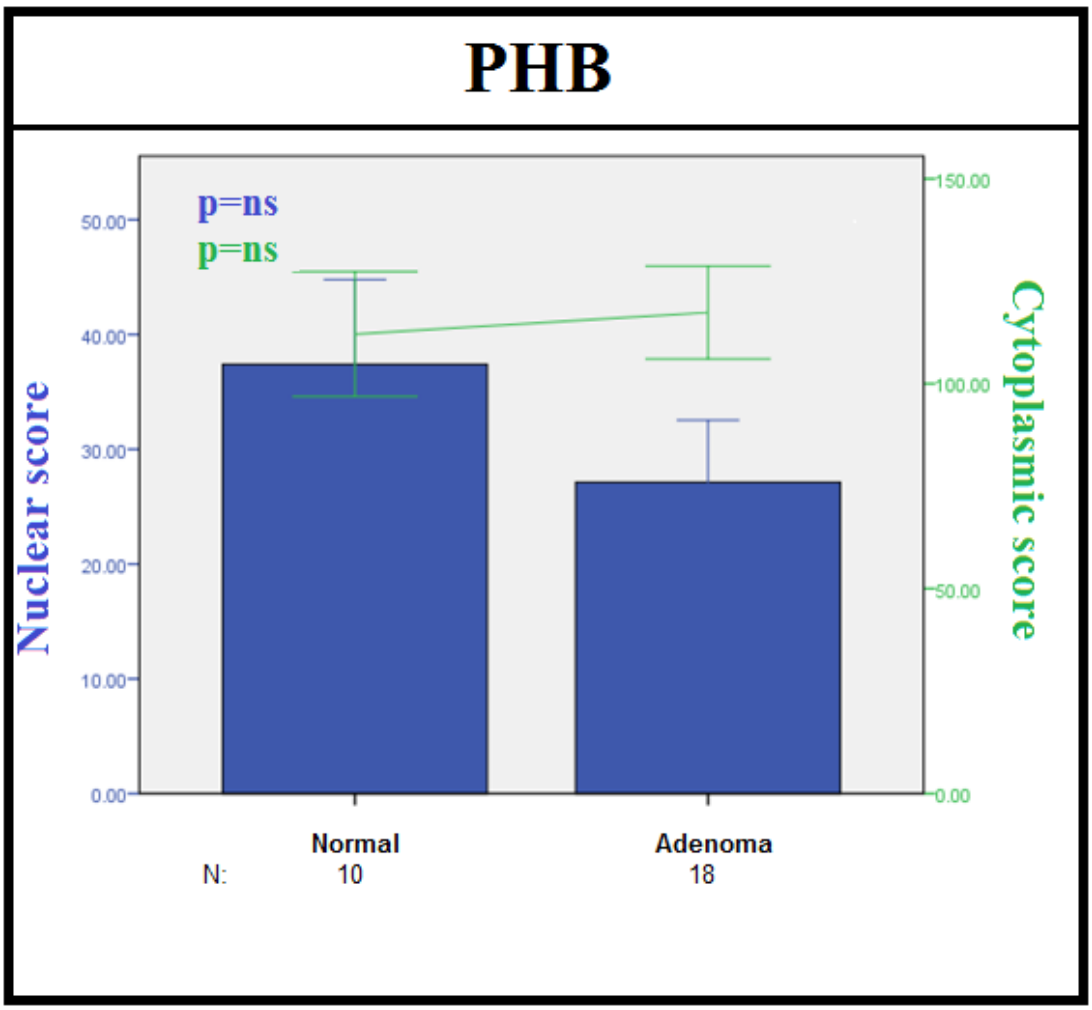


Figure 3.23. Analysis of PHB immunostaining in normal and adenoma samples from the Brazilian cohort. Again, no difference was found between normal and adenoma samples although a trend towards a decreased nuclear expression in adenoma tissues is noted (Mann-Whitney U test). Error bars represent ± 2 SE. N = sample numbers.

Prohibitin has been quite well studied in cancer models and tissues. For example, PHB immuno-expression has been shown to be reduced in ovarian cancer (Jia *et al*., 2014). Conversely, it was increased and associated with adverse prognostic features in lung (Jiang *et al*., 2013, Guo *et al*., 2012), breast (Najm *et al*., 2013), gastric (Kang *et al*., 2008), thyroid (Franzoni *et al*., 2009), prostate (Ummanni *et al*., 2008) and bladder cancers (Wu *et al*., 2007). In CRC, Chen *et al*. reported an analysis of PHB expression using IHC (Chen *et al*., 2010a). Their results were similar to ours in terms of the cytoplasmic expression of the protein. They found an increased expression in cancer tissues compared to the adjacent unaffected

mucosa. Additionally, they demonstrated that the cytoplasmic expression of PHB was similar between normal tissues and adenomas, another finding consistent with our results. They have however not reported an analysis of the nuclear expression of the protein as we have. Therefore, besides demonstrating further evidence for the differential expression of PHB in the cytoplasm of CRC cells, we provide the first insight of the nuclear staining pattern of this protein in this disease.

3.7. Expression of HMGB1 in CRC tissues

Adjacent non-neoplastic and cancerous tissues from the Brazilian cohort were sectioned and stained as per the IHC protocol described in *Chapter 2* and the images produced were analysed using the ImageJ-based electronic plugins IHC Profiler and ImmunoRatio. As figure 3.24 shows, we observed a similar pattern of staining in both adjacent normal and colorectal tumour tissues. A uniform nuclear positivity was seen in almost all assessed tissues, whereas the cytoplasmic staining was consistently weak in malignant and non-malignant samples. The results from the electronic scoring and the statistical evaluation confirmed that both tissue types had equivalent staining scores (see figure 3.25). As no difference was seen in HMGB1 expression between cancer and adjacent tissues and considering the limited availability of precious human samples, we did not proceed with further analysis of normal and polyp samples.

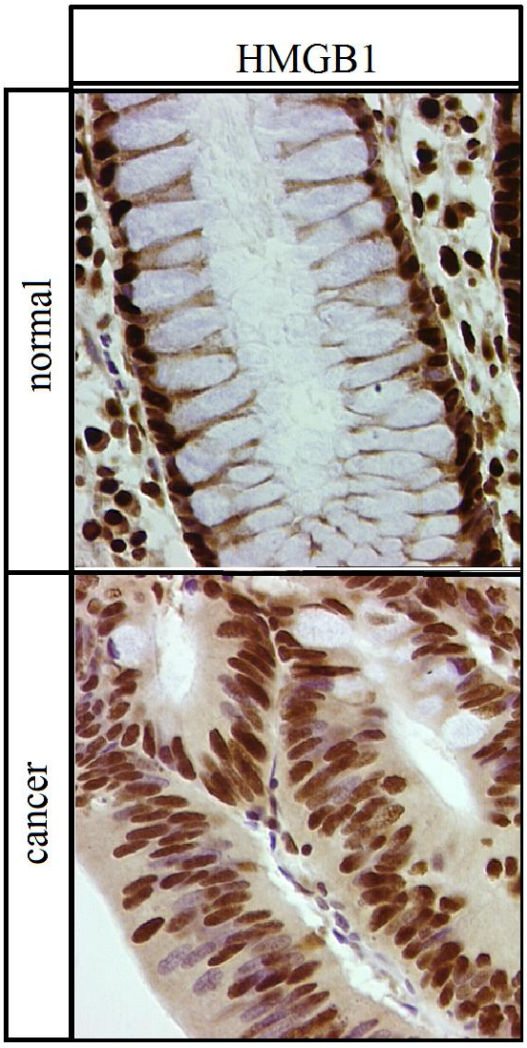


Figure 3.24. An example of HMGB1 staining pattern. We observed a similar expression of this protein in both normal and cancer tissues. Overall, the cells showed positive nuclear expression and only weak staining in the cytoplasm. These tissues were collected from a patient with stage II (Dukes' stage B) CRC. Magnification: 630x.

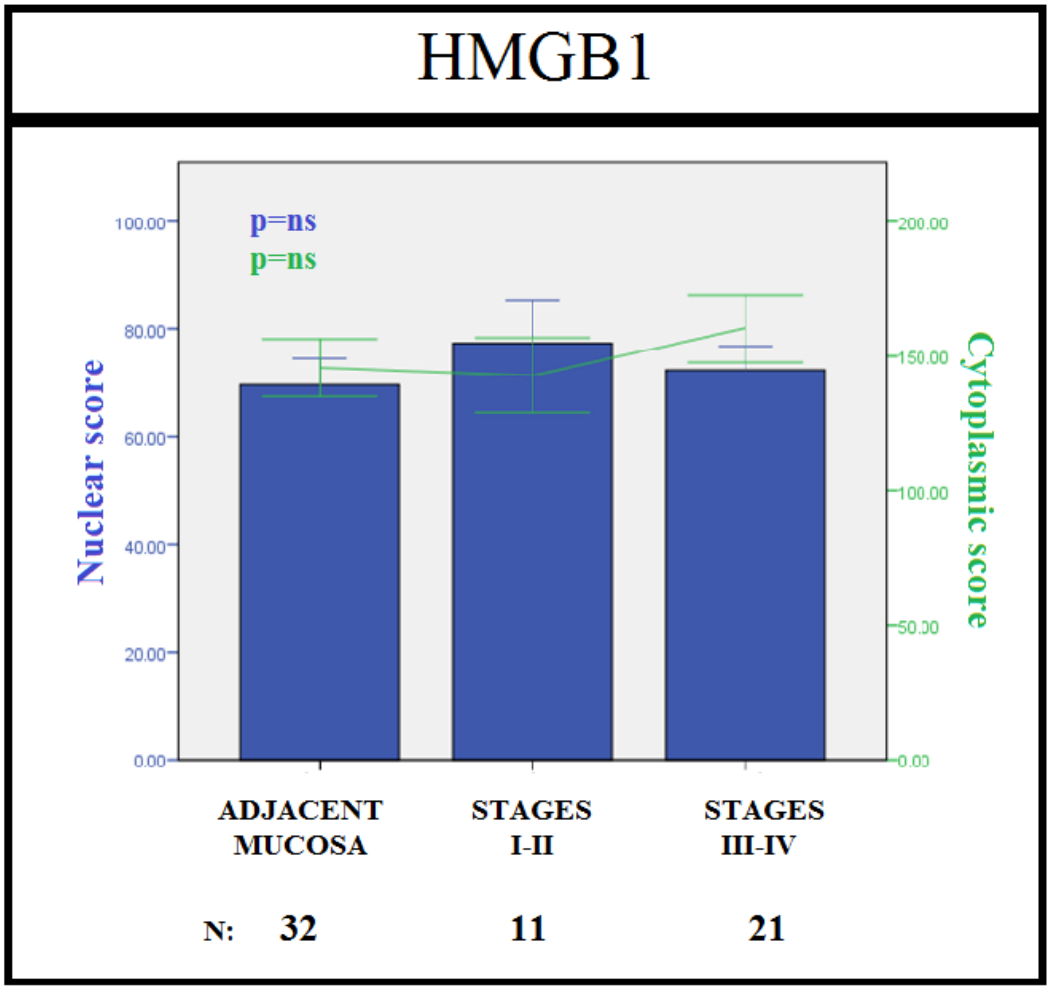


Figure 3.25. Electronic scoring of the Brazilian cohort stained for HMGB1. No difference was observed either in nuclear or in cytoplasmic staining when comparing the different groups. (Kruskal-Wallis test). Error bars represent ± 2 SE. N = sample numbers.

As mentioned in *Chapter 1 - Introduction*, HGMB1 has been extensively studied in cancer models and clinical samples including colon, breast, lung, prostate, cervical, skin, kidney, stomach, pancreatic, liver, bone and blood cancers (Kang *et al*., 2014). With reference to CRC, recent immunohistochemical evaluations have shown that cancer samples expressed higher levels of HMGB1 compared to normal adjacent tissues and this over-expression was associated with adverse prognostic factors (Süren *et al*., 2014, Zhang *et al*., 2015a, Ueda *et al*., 2014). Additionally, it was demonstrated that tumours with both nuclear and cytoplasmic expression of HMGB1 had less lymphocyte infiltration and decreased 5-year survival rates

(Peng *et al*., 2010). However, another report showed that HMGB1 nuclear immunostaining was the same in cancer and normal samples (Lee *et al*., 2012) – a finding concordant with our results.

HMGB1 has both pro- and anti-apoptotic roles in cell biology (Kang *et al*., 2013) and is involved in several disease conditions as well as cancer (Kang *et al*., 2014). This multitude of functions makes the possibility of using HMGB1 as a CRC-specific biomarker unlikely and probably explains the apparently contradictory findings in the literature. As our assessment did not confirm HMGB1's importance as a CRC biomarker in our sample set, we decided not to proceed with its evaluation in the subsequent steps of our research.

3.8. Evaluation of SFRS2 and CDC5L expression in CRC tissues

The last two candidates tested in this immunohistochemical evaluation were SFRS2 and CDC5L. Possible roles of these proteins in colorectal carcinogenesis were suggested by a previous study conducted by our group (Ibrahim, 2014). In the present analysis, we carried out IHC experiments similarly to those conducted for the previous makers using samples obtained from the Brazilian cohort of patients.

SFRS2 exhibited an almost equal pattern of nuclear/cytoplasmic balance in both normal adjacent and cancerous tissues. Homogeneous nuclear positivity was observed in malignant and non-malignant tissues. However, the intensity of nuclear staining was slightly stronger in cancer tissues than in the adjacent normal mucosa. Although difficult to perceive visually (figure 3.26), the electronic scoring produced statistically significant differences between normal and cancer groups (figure 3.27). No difference however was observed in the cytoplasmic score. Despite the statistically positive result in terms of nuclear expression, we regarded the magnitude of this difference to be small and unlikely to be clinically helpful.

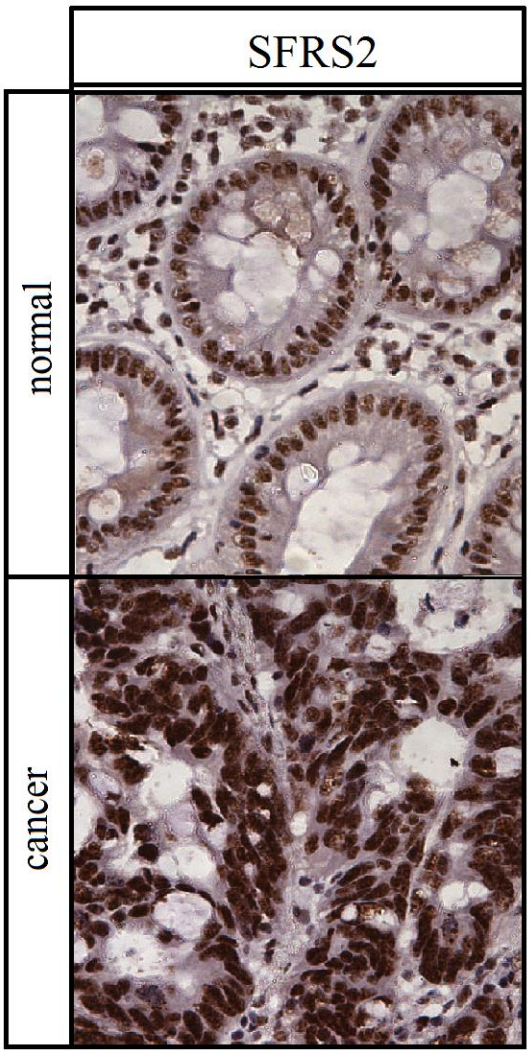


Figure 3.26. An example of SFRS2 staining pattern. Positive nuclear staining was observed in normal adjacent and cancer tissues whereas weak cytoplasmic staining was shown in both cases. These tissues were collected from a patient with stage I (Dukes' stage A) CRC. Magnification: 630x.

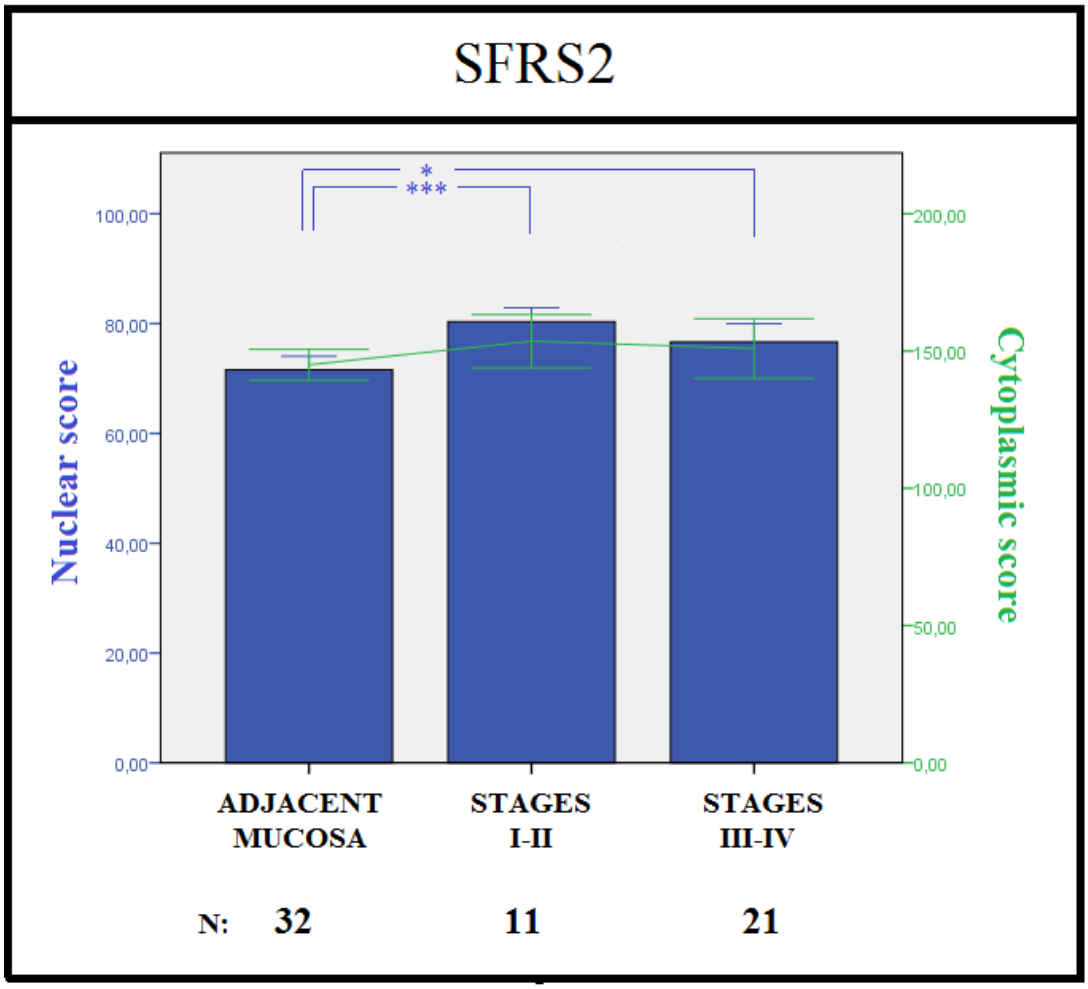


Figure 3.27. Electronic scoring of the Brazilian cohort stained for SFRS2. A small increase in the nuclear score was observed in cancer tissues compared to the adjacent mucosa. Although statistically significant, this difference was considered minor and probably clinically irrelevant (*p<0.05; ***p<0.01, Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons). No difference was seen between groups regarding the cytoplasmic scoring. Error bars represent ± 2 SE. N = sample numbers.

A similar result was observed for CDC5L. We noticed a very strong staining pattern in both adjacent normal and cancer tissues (see figure 3.28). Despite the use of the primary antibody solution in an over-diluted concentration compared with the recommended dilution (recommended: 1/250 – 1/500; used: 1/10,000), we still observed very strong DAB-staining (figure 3.28). As a result, the median nuclear scores were close to 100% in all groups. Despite this, small increases in the nuclear score were observed in the cancer groups. As the variance of the results was narrow within each group, this small difference reached statistical significance (figure 3.29). In

our opinion however, although this difference was statistically significant, it is very unlikely to have a clinical application.

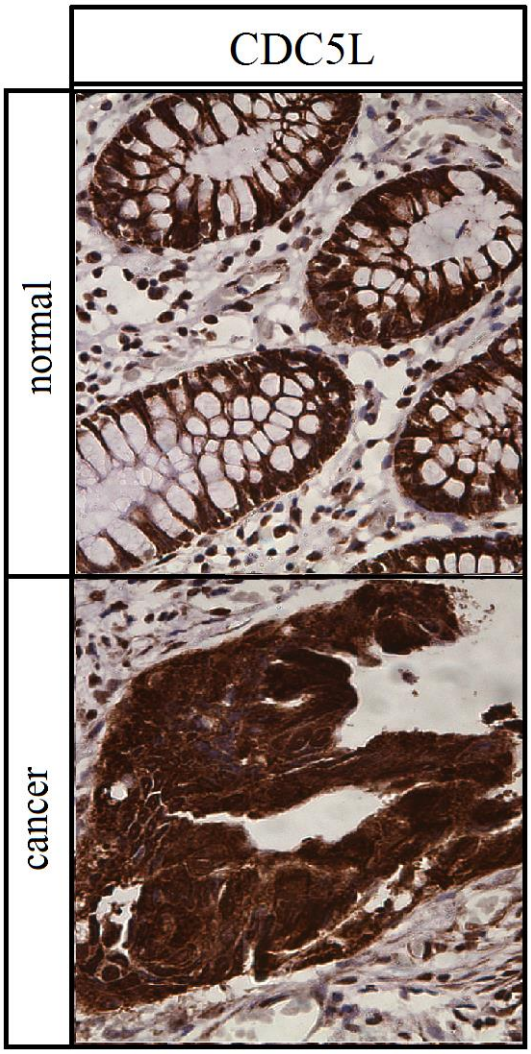


Figure 3.28. An example of CDC5L staining pattern. Despite the use of over-diluted primary antibodies, a very strong staining pattern was observed in both normal and cancer samples. These tissues were collected from a patient with stage II (Dukes' stage B) CRC. Magnification: 630x.

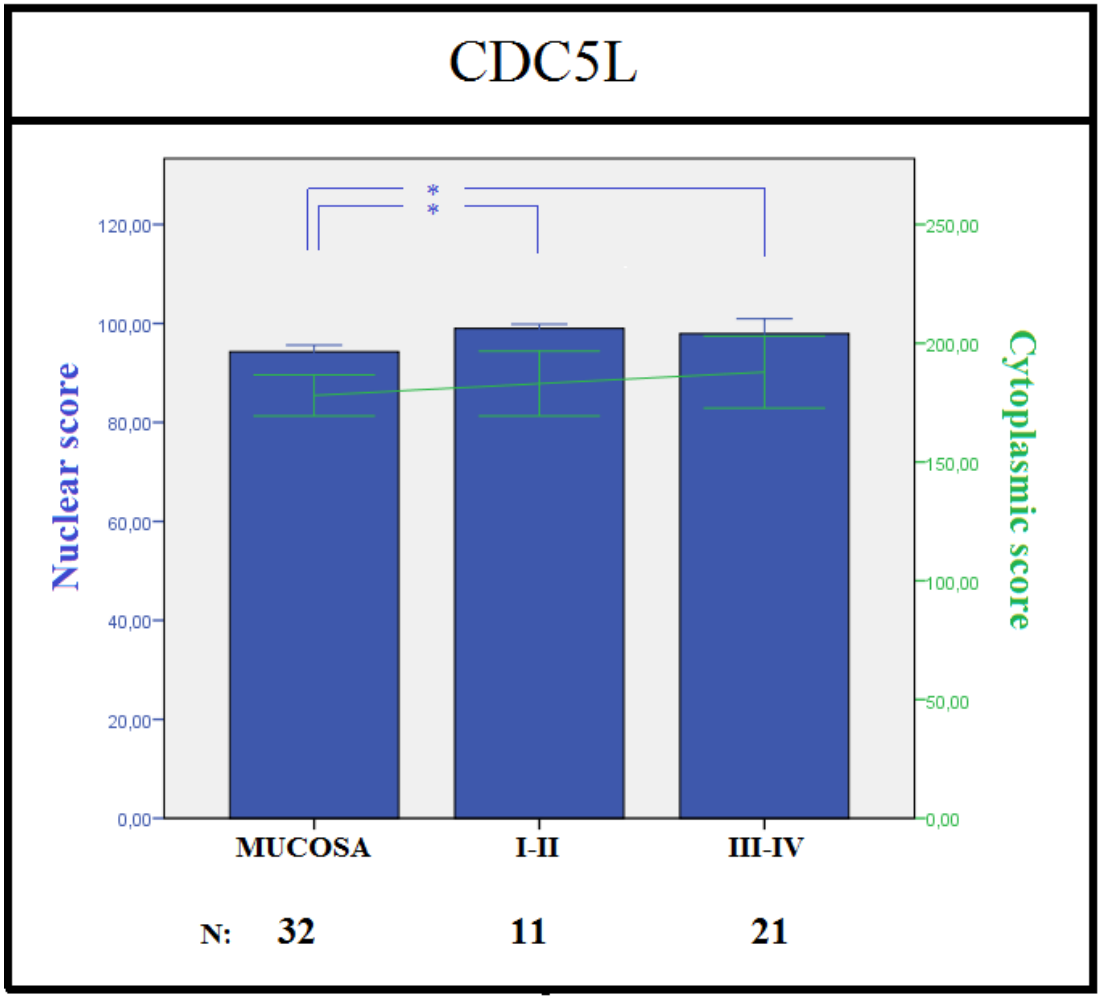


Figure 3.29. Electronic scoring of the Brazilian cohort stained for CDC5L. As observed for SFRS2, a very small but statistically significant increase in the nuclear score was observed in cancer tissues compared to the adjacent mucosa (*p<0.05, Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons). No difference was seen between groups regarding the cytoplasmic scoring. Error bars represent ± 2 SE. N = sample numbers.

We were not able to find any report of either SFRS2 or CDC5L immuno-expression in colorectal cancer tissues. Most of the studies assessing these markers have used other cancer models and cell lines, and have focused mainly upon gene expression and post-translational protein modifications (Merdzhanova *et al*., 2010, Edmond *et al*., 2011, Gräub *et al*., 2008, Mu *et al*., 2014, Lu *et al*., 2008). In this study, we did not find any relevant difference in SFRS2 or CDC5L expression that could indicate a potential role of these proteins as immunohistochemical CRC biomarkers.

Therefore, we did not investigate these two candidates further in the subsequent steps of this research.

3.9. Discussion

The discovery of novel CRC biomarkers to assist in early diagnosis, prognostic stratification and prediction of response to treatment is an urgent medical need. Immunohistochemistry has been routinely used to assess the expression of proteins with prognostic or predictive value in other types of cancer such as breast (Zaha, 2014) and lung carcinomas (Pekar-Zlotin *et al*., 2015), soft tissue sarcomas (Hornick, 2014) and lymphomas (Higgins *et al*., 2008). Besides, it is also an essential technique in the attempt to elucidate the origin of metastatic cancer from an occult primary site (Conner and Hornick, 2015, Oien and Dennis, 2012). In CRC, immunohistochemistry has mainly been used to identify the colorectal origin of metastatic lesions based of the presence CDX (an intestine specific transcription factor) and cytokeratin 20, and the absence of cytokeratin 7 (Coghlin and Murray, 2015). Mismatch repair deficiency (a surrogate marker for hereditary non-polyposis colorectal cancer) is also assessed using MLH1 and MSH2 immunostaining and it seems that the status of these mismatch-repair proteins in sporadic CRC has a prognostic role (Coghlin and Murray, 2015). Apart from these limited examples, no robust IHC biomarker has been validated for use in clinical practice. It highlights the importance of discovering new tools to better stratify cases and improve outcomes from treatment. Additionally, the definition of low-risk and high-risk adenomas is currently based on microscopic architecture, i.e. grade of dysplasia (Risio, 2010) and endoscopic features such as size, shape and number (Kurome *et al*., 2008). The discovery of an immunohistochemical profile to help accurately identify lesions with a higher risk of progression could potentially improve the management of individuals with colorectal polyps.

Mutations resulting in inactivation of APC or over-activation of β-catenin (both leading to the activation of Wnt pathway) are the earliest and

most common genetic defects in human CRC (Cancer Genome Atlas Network, 2012). Our group has hypothesised that the study of animal models of CRC based on *Apc* gene inactivation could lead to the discovery of novel useful cancer biomarkers. For this purpose, a model of acute *Apc* inactivation (*AhCre*$^{+}$ *Apc*$^{fl/fl}$ mouse) and a model of chronic *Apc* deletion (*Apc*$^{Min/+}$ mouse) have previously been studied by our group (Hammoudi *et al*., 2013, Ibrahim, 2014). This work has resulted in a list of potential candidate biomarkers which have now been subjected to a validation study using human samples.

In this study, we tested different IHC scoring systems to assess the expression of our candidate biomarkers in different pathological stages of the adenoma-carcinoma sequence and have used samples from two geographically distinct populations. Our first finding is that, independent of the scoring method used or the population assessed, the results were highly consistent. This shows that either the manual or the electronic scoring systems used are suitable for the evaluation of our candidate proteins. Additionally, it suggests that the expression pattern of the potential biomarkers is likely to be similar in different ethnic groups, as was demonstrated in this analysis using two different populations. This is important because specific genetic alterations or the expression of biomarkers may vary across different populations. For example, it has been shown that *EGFR* mutations in sporadic lung cancer are more common in Asian patients (Shigematsu *et al*., 2005) and *BRCA* germline mutations (resulting in increased risk of several types of cancer, especially breast carcinomas) are more prevalent in *Ashkenazi* Jews (Struewing *et al*., 1997).

As our candidate biomarker list was originally derived from animal models of *Apc* inactivation, our initial task was to assess Wnt signalling pathway activation in our samples. This was clearly demonstrated (in the form of β-catenin cytoplasmic and nuclear localisation) in neoplastic lesions from low-grade adenomas to invasive cancers and our findings were concordant with the literature (Wong *et al*., 2004, Wong *et al*., 2003, Chen *et al*., 2013). Next, we assessed the remaining candidates using both a manual modified-H score, and electronic nuclear and cytoplasmic scores (ImmunoRatio and IHC Profiler plugins, respectively). Results obtained for

NAP1L1 and RPL6 revealed an opposite pattern of immuno-expression when compared to β-catenin. Both proteins exhibited a clear decrease in nuclear expression in neoplastic tissues compared to the adjacent mucosa. Interestingly, the expression of NAP1L1 and RPL6 in low-grade adenomas was similar to that observed in the adjacent normal mucosa (UK-cohort) and normal control samples (Brazilian cohort), as opposed to β-catenin which exhibited an altered staining pattern even in low-grade adenomas. These findings suggest that both NAP1L1 and RPL6 might be useful immunohistochemical biomarkers for colorectal cancer and for high-grade (high-risk) adenomas. No published study assessing the immuno-expression of these two proteins in CRC tissues or adenomas has been found. Therefore, our data provide the first description of the IHC pattern of expression for both biomarkers in this context and suggest a possible role for these proteins in determining the progression from low-grade to high-grade adenomas and the subsequent development of CRC.

Another candidate that showed a clear differential expression pattern in our study was Prohibitin (PHB). As was the case for NAP1L1 and RPL6, PHB showed decreased nuclear staining in colorectal cancer tissues compared to the adjacent mucosa whereas an increase in cytoplasmic staining was observed. This interesting finding suggests that the protein might be displaced from the nucleus towards the cytoplasm during malignant transformation. Again, these alterations were specifically seen in cancer tissues and not in low-grade adenomas. Increased cytoplasmic expression of PHB in CRC has also been demonstrated in another study (Chen *et al*., 2010a). Taken together, these results support the potential use of PHB as a CRC biomarker.

HMGB1 is a multifunctional protein involved in many disease conditions (Kang *et al*., 2014). Its role in carcinogenesis has been a subject of scientific debate (Kang *et al*., 2013). In CRC tissues, HMGB1 immuno-expression has been extensively studied and mixed results have been published (Süren *et al*., 2014, Zhang *et al*., 2015a, Ueda *et al*., 2014, Lee *et al*., 2012, Peng *et al*., 2010). Our analysis performed using electronic scoring tools in cancer and adjacent tissues from the Brazilian cohort did not

demonstrate any difference in HMGB1 expression between the groups. A uniformly positive nuclear staining was observed in both tumour and adjacent tissues, whilst a negative cytoplasmic staining was seen in both conditions. These results are concordant with the findings from Peng *et al*. who demonstrated that all of the adjacent tissues in their CRC cohort exhibited positive nuclear and negative cytoplasmic HMGB1 staining. In the cancer counterpart, the same pattern was observed in 81.5% of cases (Peng *et al*., 2010).

SFRS2 and CDC5L were suggested as potential biomarkers of CRC in our prior study (Ibrahim, 2014). Neither of these proteins has been previously evaluated in CRC tissues. Our assessment of SFRS2 and CDC5L immuno-expression showed a very small increase in nuclear staining for both proteins. Although statistically significant, the small magnitude of these differences means that the findings are very unlikely to be clinically relevant.

The cause of this "switching" pattern of expression observed with NAP1L1, RPL6 and PHB (which seem to exit the nucleus during malignant transformation) is not clear. Gene expression may be altered secondary to Wnt pathway activation as occurs with *MYC* (He *et al*., 1998, Myant and Sansom, 2011) or protein localisation may be affected by post-translational modifications, as in the case of β-catenin following *APC* inactivation (Clevers and Nusse, 2012). Alternative splicing could also be involved in this setting, causing the alterations that we observed.

Altogether, our findings suggest that potential cancer biomarkers derived from animal models of *Apc* inactivation may yield valid candidates for human CRC. NAP1L1, RPL6 and Prohibitin immuno-expressions were consistently altered in malignant and high-grade premalignant colorectal neoplasms in a clear, reproducible and statistically significant manner. Whether the immuno-expression patterns of these proteins can be used as diagnostic tools or for risk stratification requires further investigation. The possibility of using these markers to identify adenomatous polyps which have a higher risk of progression to invasive cancer is particularly attractive. However, the assessment of the suitability of these proteins for this purpose

requires a larger longitudinal study and is beyond the remit of the present work. In the next chapters, I will further describe the assessment of the expression of these biomarkers using different methodologies and different matrices such as blood and frozen tissues. In addition, I will explore the prognostic importance of these proteins in CRC. As HMGB1, SFRS2 and CDC5L were not significantly altered in neoplastic colorectal tissues compared to the adjacent normal mucosa they have not been further assessed in the next steps of this research.

Chapter Four:
Study of the RNA expression of the biomarkers in human samples and the effects of gene silencing on a CRC cell line

4. CHAPTER 4 – STUDY OF THE RNA EXPRESSION OF THE BIOMARKERS IN HUMAN SAMPLES AND THE EFFECTS OF GENE SILENCING ON A CRC CELL LINE

4.1. Introduction

As mentioned earlier, work conducted in our laboratory showed that some of the candidates exhibited altered gene expression in tissues from animal models of *Apc* inactivation (Hammoudi *et al*., 2013, Ibrahim, 2014). Therefore, we decided to perform an assessment of the mRNA expression of our candidate biomarkers in human samples in order to validate the previous results. For this analysis, we used quantitative polymerase chain reaction (qPCR), as per the protocols described in *Chapter 2*.

Polymerase chain reaction (PCR) is a process that uses an *in vitro* reaction based on the activity of the enzyme *DNA polymerase* in order to amplify specific DNA sequences (Mullis *et al*., 1986, Saiki *et al*., 1988). With the capacity of producing 100 billion copies from a single sequence in a few hours, the method has been extensively used for identification of mutations, screening of pathogens, generation of forensic evidence, isolation of genes for cloning experiments, to cite a few examples (Templeton, 1992). Quantitative (also called “real-time”) PCR was first described in 1992 (Higuchi *et al*., 1992) as a method that uses fluorescent substances that bind to DNA amplicons generated during each PCR cycle. With the use of thermocyclers capable of continuously measuring the intensity of fluorescence, it permits a real-time quantification of the amplification products after each cycle.

This technology has become an ancillary technique in cancer research and clinical practice. Its applications include the detection of residual disease in leukaemia, assessment of tumour immunology, measurement of DNA copy number and detection of genomic mutations and polymorphisms (Mocellin *et al*., 2003). However, the performance of the test is highly dependent on several steps which must be carried out meticulously, from sample collection

and storage to the final reagent optimisation. The method of assessing the results is also essential for obtaining accurate results. Despite the thousands of published studies using qPCR in cancer research, many of them have flaws or omissions in the report that prevent proper validation of the findings (Bustin *et al*., 2013). This is specifically the case in CRC research. A review of the published literature has shown that 75% of the papers did not report the efficiencies of the PCR reactions, an essential parameter for assessing the validity of the results (Dijkstra *et al*., 2014). Several other omissions were also found in the same survey. As a consequence of such heterogeneity in the description of qPCR studies, a standardisation process became necessary. In 2009, an attempt to correct this issue was initiated with the publication of the *M*inimum *I*nformation for Publication of *Q*uantitative Real-Time PCR *E*xperiments (MIQE) guidelines (Bustin *et al*., 2009). This consensus proposed a list of mandatory or recommended data that should ideally be included in any report of qPCR experiment. Although the publication from Dijkstra *et al*. cited above included studies published from 2006 to 2013 (thus encompassing studies published 4 years after the publication of the recommendations), only 1% of them has cited the MIQE guidelines. In our research, we tried to follow the relevant MIQE recommendations whenever they applied.

The specific qPCR platform chosen for this research was the gene expression technology provided by Applied Biosystems – TaqMan® assays (Foster City, CA, USA), which use the so called 5´- nuclease activity of DNA polymerase. These assays include a pair of primers (forward and reverse) and a sequence-specific hydrolysis probe. At the 5´ end of the probe, a fluorescent reporter dye is attached. At the opposite 3´ end, a non-fluorescent dye (also known as “quencher”) is located. Due to the proximity of these two dyes, a phenomenon called “fluorescent resonance energy transfer” (FRET) occurs. In FRET, the emissions from the fluorescent dye are strongly reduced by the presence of a quencher. During each PCR cycle, the probe anneals with the target sequence. When the polymerase, during the extension period of the cycle, reaches the probe, the enzyme´s 5´-exonuclease activity cleaves the probe from the 5´ end, thus releasing the

reporter dye. Once free from the probe and, consequently, liberated from quencher inhibition, the reporter emits fluorescence. The fluorescence accumulates as the successive cycles progress and it is recorded by the qPCR thermocycler. TaqMan® technology was selected because it is a widely accepted system and also because Applied Biosystems´ qPCR equipment was available in the labs involved in this research in both the UK and Brazil.

The choice of a proper reference gene for the normalisation of qPCR experiments is deemed to be an essential step for accurate results. Ideally, the reference gene should maintain the same expression level in all experimental conditions. However, there is no agreement on which genes are ideal references for CRC*. ACTB* (the gene encoding β-actin*)* has been the most commonly used endogenous control to normalise qPCR experiments in CRC (Dijkstra *et al*., 2014). Nonetheless, several validation studies assessing its role as a CRC reference gene have produced inconsistent results. One report has shown that *ACTB* expression levels have a slightly higher coefficient of variation than beta-2-microglobulin (*B2M*) and peptidylprolyl-isomerise A (*PPIA*) in CRC and normal adjacent samples (Kheirelseid *et al*., 2010). Another study suggested that the use of one of the gene pairs *HPRT1/PPIA* or *IPO8/PPIA* is better than using *ACTB* alone (Sørby *et al*., 2010). Conversely, Miyata *et al*. have recently shown that *ACTB* and TATA-box binding protein (*TBP*) were stably expressed in CRC when compared with the adjacent mucosa (Miyata *et al*., 2015). In the same report, *HPRT, GAPDH, SDHA, UBC, B2M* and *18S* rRNA all exhibited altered expression. Given the lack of a standard reference gene in this context, we decided to use *ACTB* as our endogenous control. Besides our candidate biomarkers (*NAP1L1, RPL6 and PHB*), we also assessed β-catenin gene (*CTNNB1*) expression to have a better understanding of the status of Wnt signalling pathway in these cases.

The functions of particular genes can be explored by interfering with gene activity. Gene expression stimulation or silencing may provide important clues regarding its role in cell physiology. Given the paucity of information concerning our candidate biomarkers in terms of cell biology regulation, RNA

interference studies were performed targeting the candidates which have exhibited significant gene expression alterations in the human tissue analysis explained above. Details of these experiments are described later in this chapter.

4.2. Clinical samples

This study involved exclusively samples collected in Brazil. Cancer and adjacent samples were collected during the surgical procedure immediately after the removal of the specimen, thus avoiding long term tissue hypoxia. Samples were immediately immersed in Allprotect® (Qiagen), incubated overnight at 4°C and stored at -80°C until they were used. Normal control samples were collected from individuals who had no endoscopic abnormality during colonoscopy. These samples were also incubated in Allprotect® and stored similarly to the cancer samples. RNA extraction and cDNA synthesis were processed as described in *Chapter 2*. The clinical and demographic characteristics of the individuals included in this analysis are summarised in table 4.1. As can be noticed, the groups were well matched in terms of age. However, there was a predominance of males in the cancer group whilst females were the majority in the normal group. It has been demonstrated that women are more likely to seek medical attention for health promotion and to participate in screening programmes than men (Sach and Whynes, 2009). This may be even more relevant in regions without an organised programme such as Brazil and may have contributed to the predominance of women in the normal control cohort.

Table 4.1. Clinical and demographic characteristics of the individuals included in the qPCR analysis.

Characteristics	**Normal control (n=10)**	**Cancer (n=25)**	**p**
Mean age (range)	54.8 (41-74)	55.9 (34-83)	NS
Gender			0.001*
Male	2	16	
Female	8	9	
Stage			NA
I-II	-	42%	-
III-IV	-	58%	-

* Fisher´s exact test. NS: non-significant. NA: not applicable.

4.3. Assessment of qPCR assay efficiencies

When using commercially available qPCR assays, many researchers assume that the kits have very high efficiencies. However, as recommended by the MIQE guidelines, this must be confirmed to be the case in the experimental conditions of each PCR lab. Additionally, the manufacturer of the kit that we used also recommends that the efficiency be tested before the use of the assays (*Relative Gene Expression Workflow document*, available at http://tools.thermofisher.com/content/sfs/brochures/cms_075428.pdf, last accessed on 29/12/15). Checking the PCR efficiency is essential, as the most commonly used method for analysing the results, the "comparative Ct" or "$2^{-\Delta\Delta CT}$" ("Livak") method is only suitable for experiments in which the efficiencies of the amplification for the endogenous control and the target gene are similar and close to 100%. The developers of the method recommend that the efficiencies do not differ by more than 10%, and lie between 1.8 and 2.2 (equivalent to 90 to 110% of efficiency) (Schmittgen and Livak, 2008).

In this research, we followed the recommendation provide by the manufacturer of the kit (described above). The assessment of PCR

efficiencies was performed by the amplification of a dilution series (10-fold dilutions, 5 points) using cDNA from one control sample. This process was repeated for each gene (reference gene and targets). The threshold cycle, or Ct, was annotated for each dilution point and a regression curve was plotted. Using the ABI 7500 real-time PCR system software, the slope of the curve was calculated and the efficiency of the reaction was obtained. Figure 4.1 shows the result produced for *CTNNB1*. The assay exhibited an efficiency of 93.6%. Efficiency results *for ACTB, NAP1L1, RPL6* and *PHB* ranged from 90.0% to 96.7%. Therefore, the assumptions of both similar and high efficiencies were satisfied for all genes and the comparative Ct method was deemed adequate for our analyses.

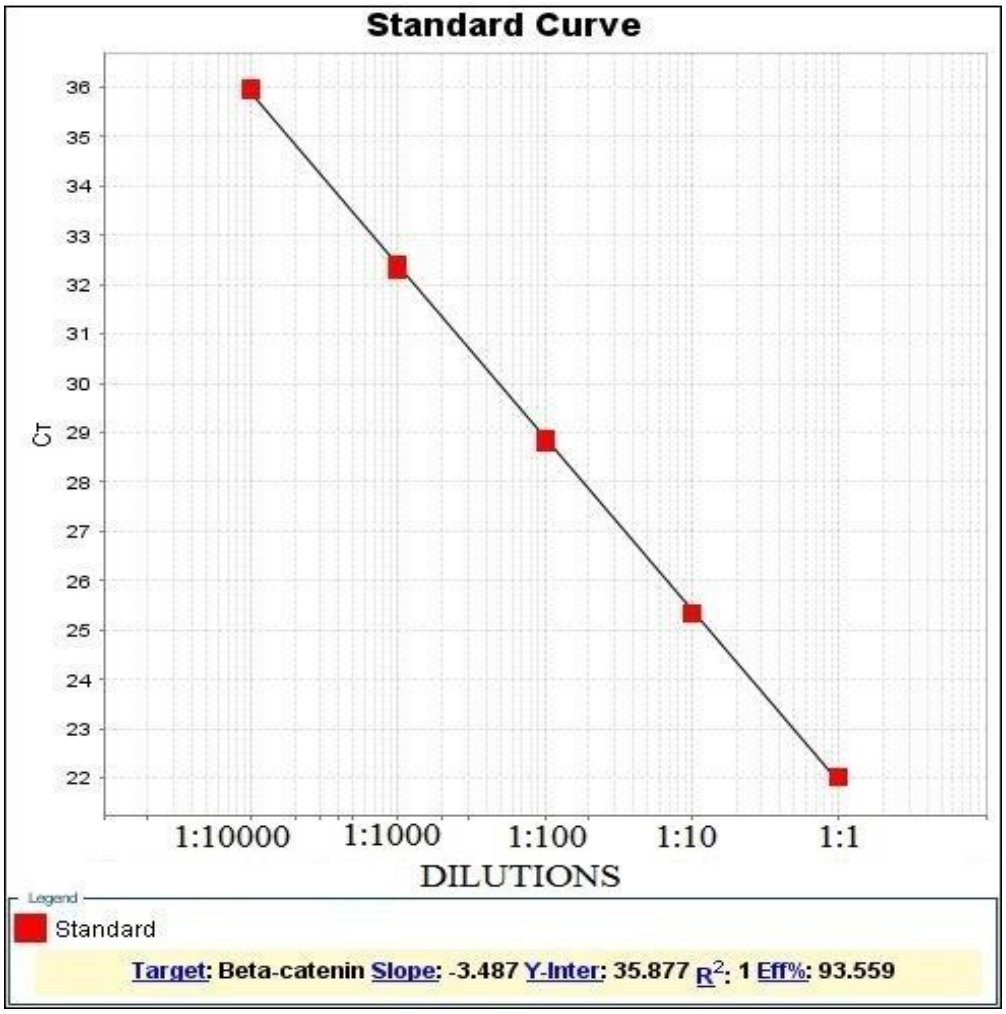


Figure 4.1. Graph representing the assessment of PCR efficiency for *CTNNB1*. A 10-fold dilution curve (5 points) was produced using cDNA from one sample and the dilutions underwent amplification in triplicate. A linear curve was obtained, showing that the dilutions were properly prepared. In this case, an efficiency of 93.6% was found. Similar procedures were performed for the other genes.

4.4. Defining the best cDNA dilution for the qPCR experiments

In *Chapter 2*, it was explained that the amount of RNA used in the reverse transcription step for cDNA synthesis may affect the efficiency of the reaction mainly due to contamination carry-over and internal inhibitors (Bustin *et al*., 2015, Pugniere *et al*., 2011). The same notion applies to the qPCR step. RT reagents can work as PCR inhibitors, thus affecting the performance of the final reaction (Gallup and Ackermann, 2006). To address this issue, most gene expression researchers dilute the cDNA from 1:2 to 1:100, and use the diluted stock solution to carry out qPCR experiments. However, when working with genes whose level of expression is not well known, it is highly advised that different cDNA dilutions be tested in order to choose an appropriate concentration. This step aims to avoid the use of excessively concentrated post RT solution (thus preventing any contamination carry-over) and to reassure that the Ct for each gene is located between cycles 15 and 30, a range in which the qPCR reaction is more accurate.

In our experiments, 300ng of RNA were used in the RT step after testing different quantities, as described previously. In order to find an adequate cDNA dilution, we performed amplification using a dilution series (2-fold, 5 points) ranging from 1:1 to 1:16 (one control sample). Experiments were repeated for all genes, both targets and reference. For each gene tested, all dilutions resulted in Cts between 15 and 30 (as an example, see the amplification plot for *CTNNB1* in figure 4.2). Given these results, we could confidently choose any of the dilutions tested. To avoid using unnecessary amounts of precious samples, but also preventing any particularly low expressing sample from exhibiting late amplification, we decided to use the 1:4 dilution in these experiments.

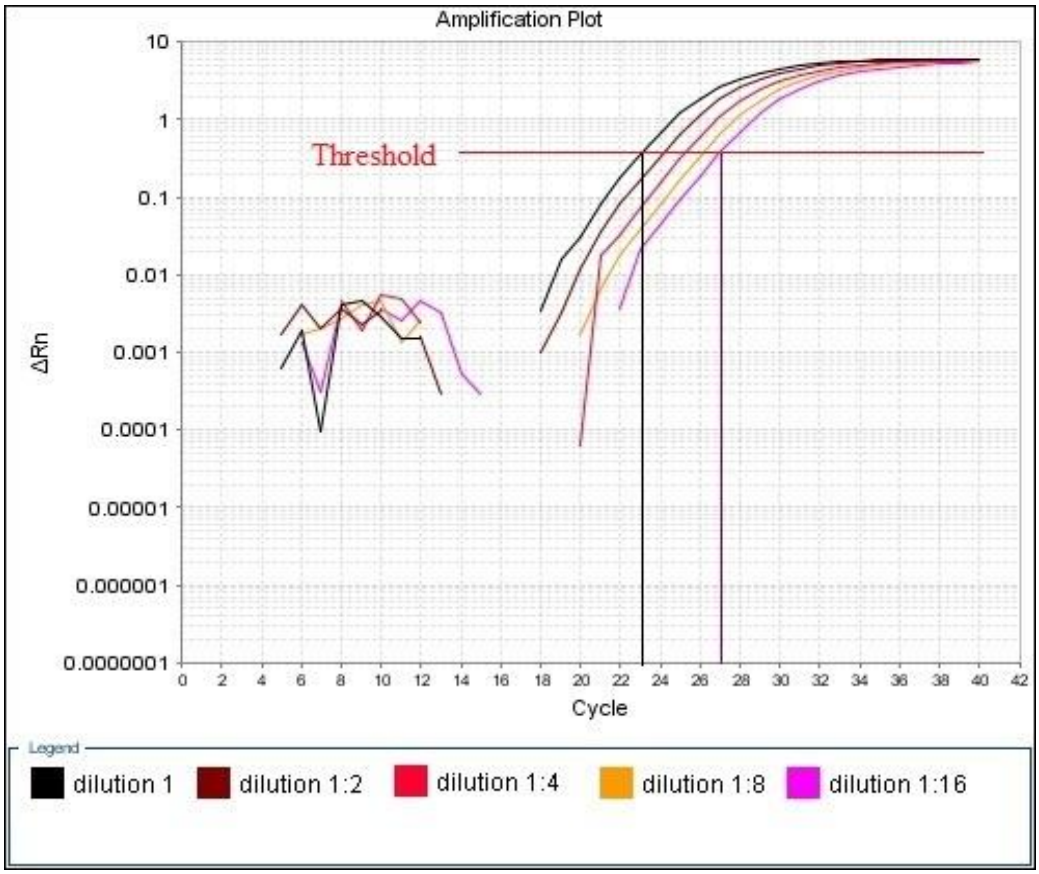


Figure 4.2. Assessment of different cDNA template dilutions using *CTNNB1* assay. The serial dilution produced distinct and ordered amplification curves, with all Cts lying between cycles 15 and 30. Similar results were obtained for the other genes.

Although most publications in the literature describe gene expression studies comparing only tumour samples and adjacent normal epithelium, we decided to include in our analysis normal samples which had been obtained from individuals without any endoscopy-detectable neoplasm or inflammatory condition. The reason for this is the evidence in the literature of altered gene expression in the apparently normal mucosa adjacent to epithelial tumours (Raudenska et al., 2015a, Sanz-Pamplona et al., 2014, Chandran et al., 2005), a phenomenon known as "*field cancerisation*" or "*field effect*". Consequently, a more comprehensive understanding of the role of the candidate genes in both the tumour and the surrounding environment could be obtained. In the next sections, the results from the assessment of the expression of the candidate genes in human tissues will be presented.

4.5. Expression of the candidate genes in human tissues

4.5.1. Expression of *CTNNB1* in human tissues

In addition to the immunohistochemical expression/localisation of β-catenin work detailed in *Chapter 3*, we also assessed *CTNNB1* mRNA expression in our sample cohort. For this purpose, cDNA from normal control, adjacent unaffected colon and cancer tissues was analysed by qPCR according to the protocols described in *Chapter 2*. Each reaction was carried out in triplicate and non-template controls were used in each PCR plate. As shown in figure 4.3, we observed similar expression levels comparing normal controls and adjacent non-neoplastic samples. However, a small but significant increase in mRNA expression was observed in cancer samples compared to adjacent unaffected colon samples (but not to normal controls).

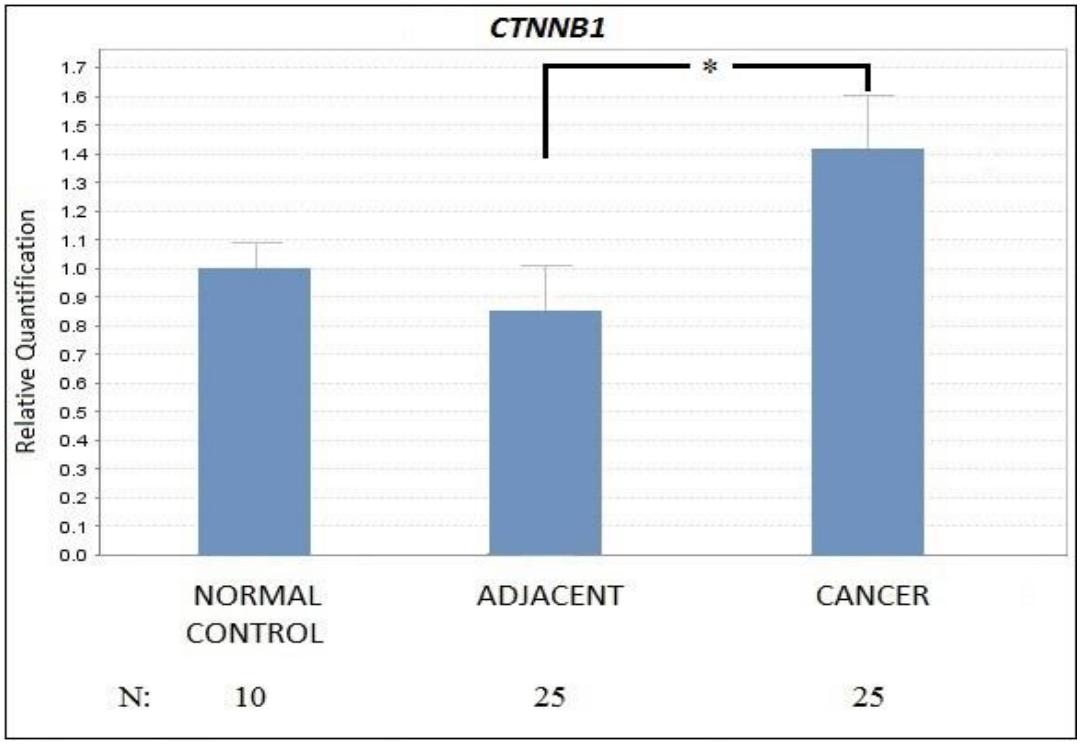


Figure 4.3. *CTNNB1* mRNA expression in normal, adjacent and cancer samples. When normalised to *β-actin*, *CTNNB1* exhibited stable levels in normal (RQ 1.0; min-max 0.92-1.09) and adjacent (RQ 0.85; min-max 0.71-1.01) samples. In cancer samples, *CTNNB1* expression was slightly increased (RQ 1.42; min-max 1.26–1.60). Although small, this expression was significantly higher compared to the adjacent but not to the normal group. * $p<0.05$ (Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons).

The literature is rich in reports of the immunohistochemical expression of β-catenin in several cancer types. However, few studies have analysed β-catenin gene expression either in CRC or in other cancers. The limited number of results that we found has shown contradictory findings. Anwar *et al*. demonstrated a 21-fold increase in *CTNNB1* mRNA expression in CRC tissues compared to both the "adjoining" (2-5cm from the tumour) and "normal adjacent" mucosa (5-10cm from the tumour) (Anwar *et al*., 2015). Conversely, a Chinese report showed a decreased expression of *CTNNB1* in cancer tissues when compared to the adjacent epithelium (Qin *et al*., 2006). Another study found no difference in *CTNNB1* expression when comparing tumours and adjacent tissues (Truant *et al*., 2008). Our results support the concept that *CTNNB1* activation is not a hallmark of CRC as only a minor increase in expression was observed in cancer tissues compared to adjacent tissues. The difference between normal and cancer tissues was not statistically significant. Additionally, no difference was seen when comparing normal to adjacent tissues.

4.5.2. **Expression of *NAP1L1* in human tissues**

NAP1L1 mRNA expression was assessed in CRC, adjacent mucosa and normal controls using a TaqMan® gene expression assay. Before analysing the samples, PCR efficiency was assessed and the optimal amount of RNA template was determined, as detailed above. Then, the experiment was carried out and the results were compared using *ACTB* as the reference gene and the comparative Ct method. There was an evident and statistically significant increase in *NAP1L1* expression in both tumours (8.28-fold) and adjacent tissues (7.18-fold) when compared to normal controls, as shown in figure 4.4. No difference between tumours and adjacent tissues was observed.

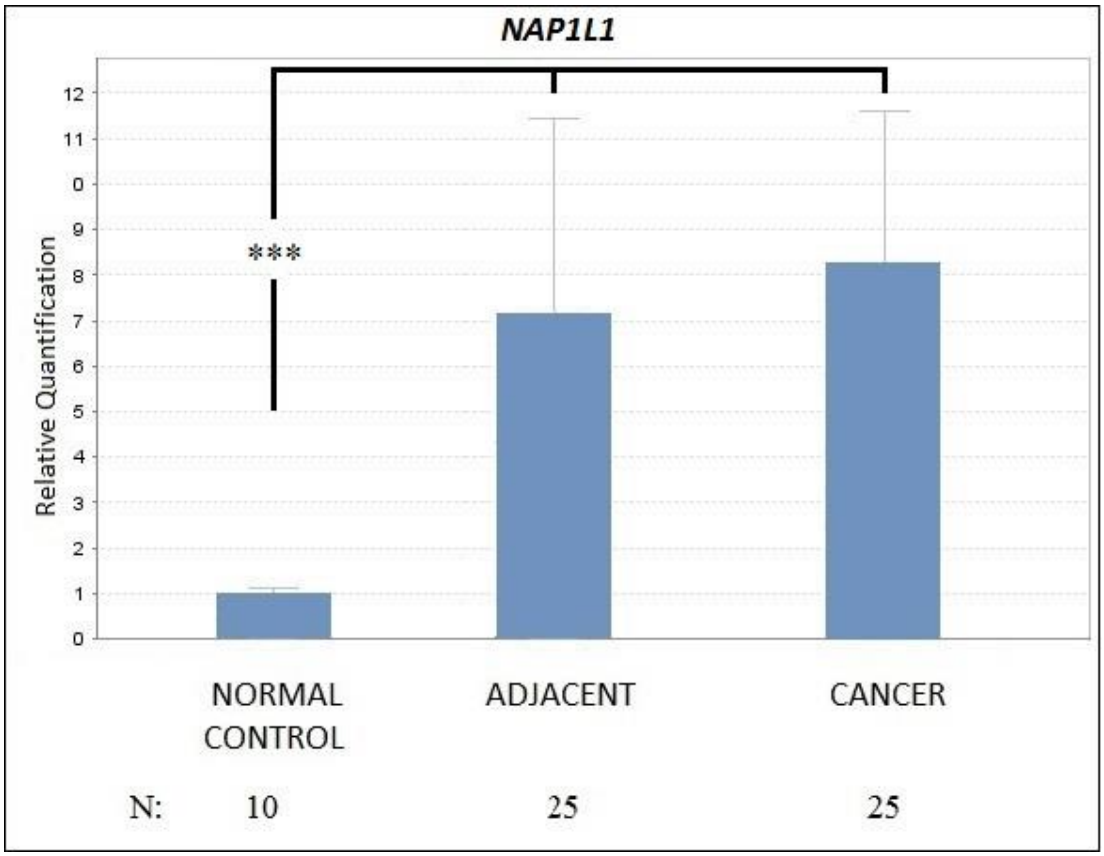


Figure 4.4. *NAP1L1* mRNA expression in normal, adjacent and cancer samples. A major increase in *NAP1L1* expression was observed in both adjacent (RQ 7.18; min-max 4.50-11.44) and tumour (RQ 8.28; min-max 5.92-11.60) tissues compared to normal controls (RQ 1.00; min-max 0.89–1.13). *** $p<0.001$ (Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons).

Few researchers have studied *NAP1L1* expression in cancer tissues. Drozdov *et al*. compared small intestinal neuroendocrine tumours (NETs) and normal enterochromaffin cell preparations, and showed a 13.7-fold increase in *NAP1L1* expression in tumour tissues (Drozdov *et al*., 2009). However, no analysis of the adjacent mucosa was performed. Studying both NETs and CRCs, Kidd *et al*. suggested that *NAP1L1* was increased in NETs but not in CRCs (Kidd *et al*., 2006). However, the comparisons were made with the respective adjacent mucosas and no tissue from healthy individuals was tested. Therefore, their results are not discordant when compared with ours. Line *et al*. evaluated *NAP1L1* mRNA expression in CRC and adjacent tissues as a secondary endpoint in a study primarily aimed at finding sero-reactive biomarkers (Line *et al*., 2002). They showed that, among 15 cases of CRC, seven exhibited moderate increases in *NAP1L1* expression (ranging from 2.9

to 9.3-fold) and eight cases showed expression levels similar to the corresponding adjacent mucosa. Again, the fact that they used only adjacent tissues may have prevented them from finding a more pronounced differential expression. To the best of our knowledge, these results show for the first time a differential expression of *NAP1L1* in CRC and adjacent tissues, suggesting a possible role for this candidate biomarker not only in malignant transformation but also in the process of field cancerisation.

The precise interpretation of these findings in comparison with the results of the assessment of NAP1L1 immunohistochemical expression (*Chapter 3*) requires further research. We previously observed an evident decrease in the nuclear immuno-expression of NAP1L1 in cancer tissues when compared with the adjacent unaffected colonic mucosa. As we now demonstrate that the mRNA expression of this marker is similarly increased in both samples – tumour and adjacent, it is likely that another event may have occurred during the process of malignant transformation, affecting the protein content and distribution within the cell. Either increased protein degradation or post-translational modifications might have caused the decreased immunostaining observed. Alternatively, the protein might have left the cell via excretion, secretion or leakage. In the study mentioned above, Line *et al*. showed that NAP1L1 caused an immunogenic response in a subset of CRC patients but not in healthy volunteers, thus supporting the notion that the protein may leave the cell upon cancer development (Line *et al*., 2002).

4.5.3. **Expression of *RPL6* in human tissues**

Similar experiments were performed to assess *RPL6* gene expression. Using TaqMan pre-optimised assays, we compared the transcript levels of this candidate biomarker in tumours, adjacent tissues and normal controls. In the same way that we observed in the immunohistochemical work, *RPL6* gene expression mirrored the pattern observed for *NAP1L1*. As depicted in figure 4.5, an increase in *RPL6* expression was found in both adjacent and

tumour tissues when compared to normal controls. Although this difference was statistically significant, the magnitude of the over-expression was less than that observed for *NAP1L1*. Again, no difference was observed between tumours and adjacent tissues.

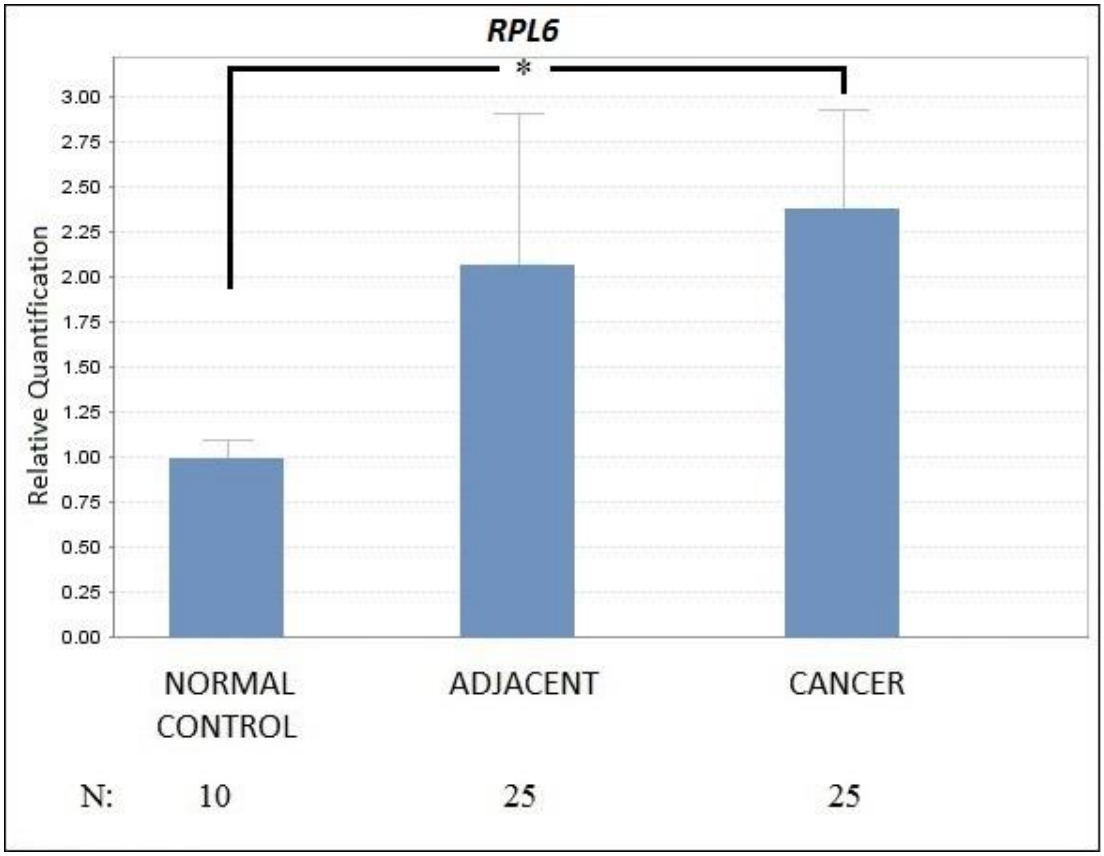


Figure 4.5. *RPL6* mRNA expression in normal, adjacent and cancer samples. Similarly to *NAP1L1, RPL6* expression was significantly increased in both adjacent (RQ 2.08; min-max 1.48-2.91) and tumour (RQ 2.38; min-max 1.94-2.92) tissues compared to normal controls (RQ 1.00; min-max 0.91–1.10). The magnitude of the difference, however, was less prominent. * $p<0.05$ (Kruskal-Wallis test followed by post hoc Dunn-Bonferroni test for pair-wise comparisons).

RPL6 expression and function has previously been studied mainly in gastric carcinomas. A Chinese group has demonstrated that this protein has important roles in cell proliferation (Gou *et al*., 2010, Wu *et al*., 2011) and drug resistance in gastric cancer cell lines (Du *et al*., 2005). Regarding CRC, the only study that we have found was the report from our group by Hammoudi *et al*. In this study, a slight increase (1.88-fold) in *RPL6* mRNA

expression was observed in tumours compared to the adjacent mucosa in a cohort of 15 patients (Hammoudi *et al*., 2013). The present study reinforces the results from Hammoudi, showing a more evident increase in *RPL6* expression in tumours compared with normal controls. Additionally, as occurred with *NAP1L1, RPL6* was shown to be over-expressed in the adjacent mucosa as well, bringing up the possibility that this gene might play a role in field cancerisation.

The same comments made in relation to NAP1L1 immunohistochemical and mRNA expressions also apply to RPL6. The apparent discordant results (decreased nuclear IHC expression and increased gene expression) suggest that some post-translational event might have occurred affecting RPL6 content and distribution. In fact, the clear correlation between the patterns of expression observed using two different techniques raises the possibility that these proteins are important and possibly related to each other in the process of colorectal carcinogenesis.

4.5.4. **Expression of *PHB* in human tissues**

The assessment of *PHB* expression was performed in the same manner described for the previous markers, using the conditions shown above and in *Chapter 2*. The comparative Ct method revealed a continuous increase in *PHB* expression from normal controls to adjacent tissues and to cancer samples. However, these differences were not statistically significant (figure 4.6).

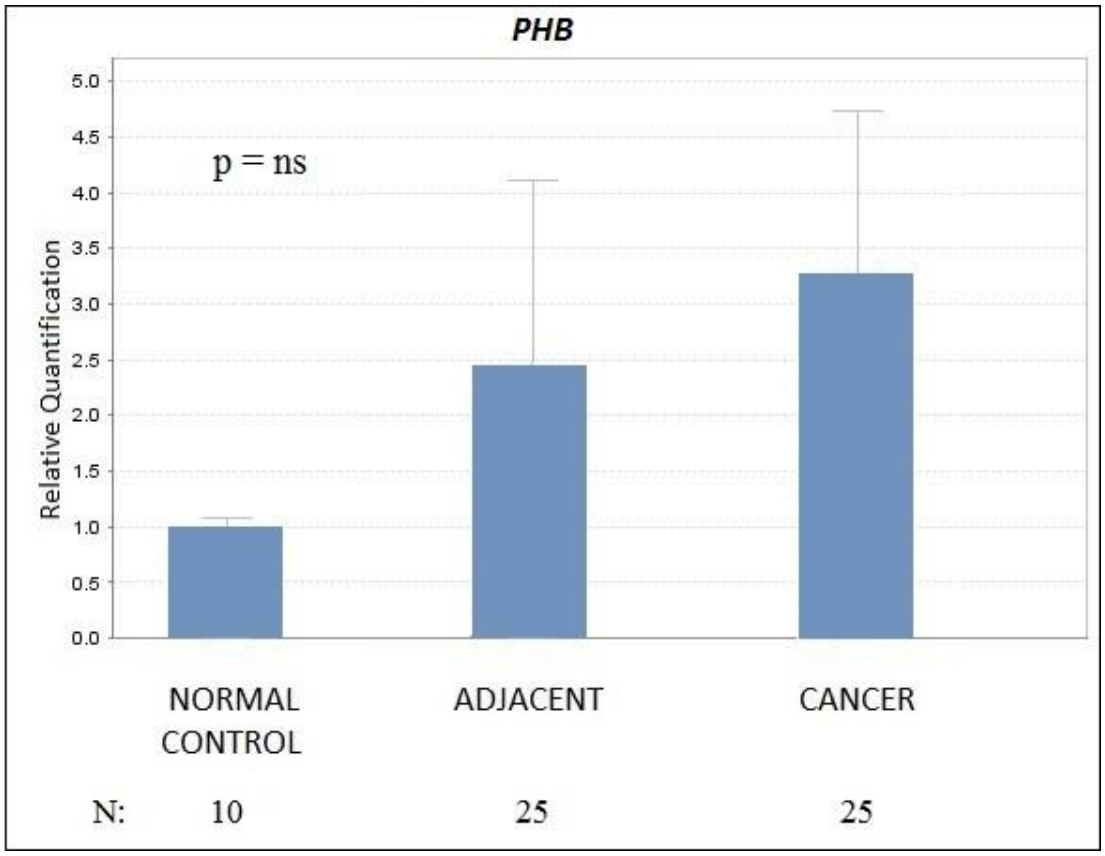


Figure 4.6. *PHB* mRNA expression in normal, adjacent and cancer samples. Despite the apparent higher expression in adjacent and cancer tissues, no statistically significant difference was observed between groups. Relative quantifications were 1.00 (min-max 0.93–1.08), 2.46 (min-max 1.48-4.11), and 3.29 (min-max 2.29-4.73) for normal controls, adjacent samples and tumours, respectively (p=0.089; Kruskal-Wallis test).

PHB expression has been studied in various cancer types using either immunohistochemistry (Guo *et al*., 2012, Jia *et al*., 2014, Wu *et al*., 2007) or combined techniques including qPCR (Franzoni *et al*., 2009, Jiang *et al*., 2013, Kang *et al*., 2008, Ummanni *et al*., 2008). All these studies have shown differential expression in cancer tissues compared to normal controls or adjacent tissues. In CRC, Chen *et al*. used comparative proteomics to screen for potential biomarkers. Using matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF/TOF-MS), they found an over-expression of PHB in cancer samples when compared to the adjacent mucosa (Chen *et al*., 2010a). Furthermore, IHC and Western blot confirmed protein accumulation in malignant tissues, a finding concordant with our

results presented in the previous chapter. Transcript levels, however, were not assessed in that study.

In our analysis, a trend pointing towards an increased expression of *PHB* mRNA in cancer tissues was seen. Nonetheless, statistical significance was not reached. This is probably due to the relatively small number of cases per group and the wide variation in the relative quantification ratios in the adjacent and cancer groups. The only report of *PHB* gene expression in CRC that we were able to find was, again, Hammoudi´s study in which the authors reported an 1.32-fold increase in CRC when compared to the adjacent mucosa (p=0.020) (Hammoudi *et al*., 2013). Along with the findings of the present research, it suggests that *PHB* transcript levels might be only marginally increased in CRC. Therefore, another mechanism different from gene over-expression might be responsible for the increased protein levels observed in our study (previous chapter) and also reported by Chen *et al*. Impaired protein degradation and post-translational modifications may explain these findings.

4.6. RNA interference studies

RNA interference is the process of gene expression silencing induced by short double-stranded RNA fragments such as small interfering RNAs or microRNAs (Rana, 2007). Small interfering RNAs (siRNAs) are produced when long double-stranded RNA is cleaved by Dicer, a type III RNAse enzyme, into fragments 21-23 nucleotides long. Each one of these fragments is incorporated into a protein complex called RNA-induced silencing complex (RISC). The antisense-, also called guide-strand, serves as a template for the identification of the target sequence. When the target mRNA is recognised by, and bound to RISC, translational repression or transcript cleavage occurs, as depicted in figure 4.7 (Rana, 2007). The technique for artificially inducing RNA interference using siRNAs was first introduced by Fire *et al*. in 1998 (Fire *et al*., 1998). Since then, the use of siRNAs has shown a massive growth in several research fields, from basic investigations to the discovery of potential gene therapies (Pecot *et al*., 2011).

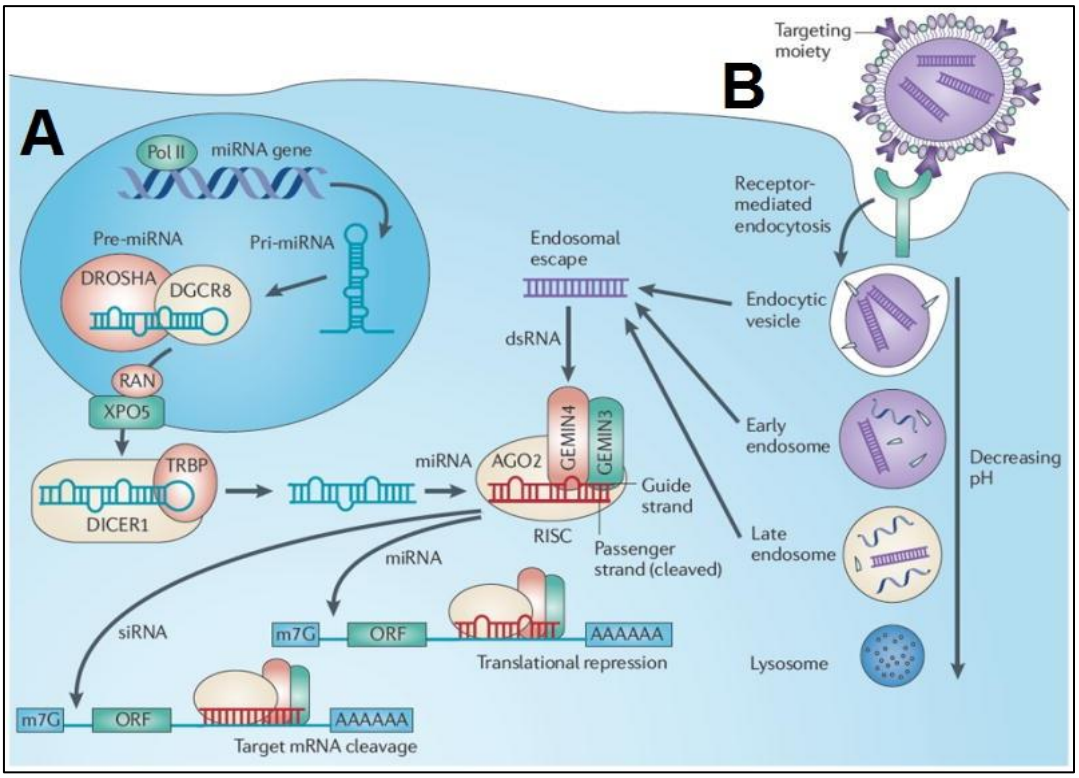


Figure 4.7. Mechanisms of RNA interference. Micro-RNAs are endogenously produced from pre-miRNAs cleaved by Dicer (A). Alternatively, siRNAs are exogenously delivered to the cell, undergo endocytosis and become free in the cytoplasm after escaping from the endosome (B). In either case, the short RNA fragment is incorporated into RISC. The guide strand provides the template for mRNA recognition. Once bound to RISC, the transcript undergoes translational repression or cleavage. From (Pecot *et al*., 2011).

Elevated levels of *NAP1L1* and *RPL6* were observed in CRC and adjacent tissues when compared with normal control mucosa, as detailed earlier in this chapter. It has also recently been shown that *NAP1L1* over-expression induces cell proliferation in murine stem cells (Yan *et al*., 2016). Other authors have demonstrated that the silencing of this gene caused immature P19CL6 cells (Li *et al*., 2012) and pluripotent stem cells (Gong *et al*., 2014), both murine cell lines, to differentiate into mature cardiomiocytes. As for *RPL6*, a study of the effect of gene knockdown using siRNA showed a negative impact on proliferation and cell cycle progression in human gastric cancer cell lines (Wu *et al*., 2011). Collectively, these findings suggest that these genes are promoters of proliferation and repressors of differentiation, two key events during the carcinogenesis process. Therefore, we hypothesised that these genes could also be important for proliferation in

human CRC cell lines. To test this hypothesis, HCT116 cells were selected and RNA interference experiments using siRNAs were performed. CRC cell lines have been widely used in studies assessing tumour biology, drug response and biomarkers. Several commonly used cell types are considered good representatives of primary colorectal tumours (Mouradov *et al*., 2014). HCT116 cells are derived from a colonic adenocarcinoma removed from a male patient. This cell line has a constitutively active Wnt signalling pathway caused by a mutation in *CTNNB1* (a three-base deletion in codon 45, exon 3), whilst the *APC* gene is wild-type (Ilyas *et al*., 1997, Morin *et al*., 1997). To assess the effect of *TP53* status on gene function, both *TP53* wild-type and *TP53* null HCT116 cells were tested.

The reagents and conditions used in these experiments were optimised by another researcher in our group prior to this study (Ibrahim, 2014). All the procedures undertaken have been detailed in *Chapter 2*. After gene knockdown, a proliferation assay was performed in order to investigate whether or not *NAP1L1* and *RPL6* regulate proliferation in this cell line. If a positive association was found, RNA extracted from transfected cells was used in a CRC developmental qPCR array plate to assess other CRC-related genes that were potentially affected by the knockdown of our markers.

4.6.1. **Assessing the efficiency of gene silencing**

HCT116 cells (*TP53* wild-type and null) were cultured and interference experiments using siRNAs were performed as detailed in *Chapter 2*. Before analysing the results, an initial mandatory step was the assessment of silencing efficiency. To be successful, a siRNA experiment must be able to significantly reduce gene expression when compared to non-transfected cells or cells transfected with non-targeting (scrambled) siRNA. Therefore, the first task in this part of the investigation was the assessment of the relative reduction in mRNA levels after the knockdown of *NAP1L1* and *RPL6*. For this purpose, cells were harvested 48 hours after siRNA transfection, RNA was extracted and cDNA was produced. Figure 4.8 shows the results for *TP53*

wild type cells. RNA interference resulted in a profound decrease in mRNA levels of both *NAP1L1* and *RPL6*, when compared to non-transfected cells or cells transfected with non-targeting siRNA.

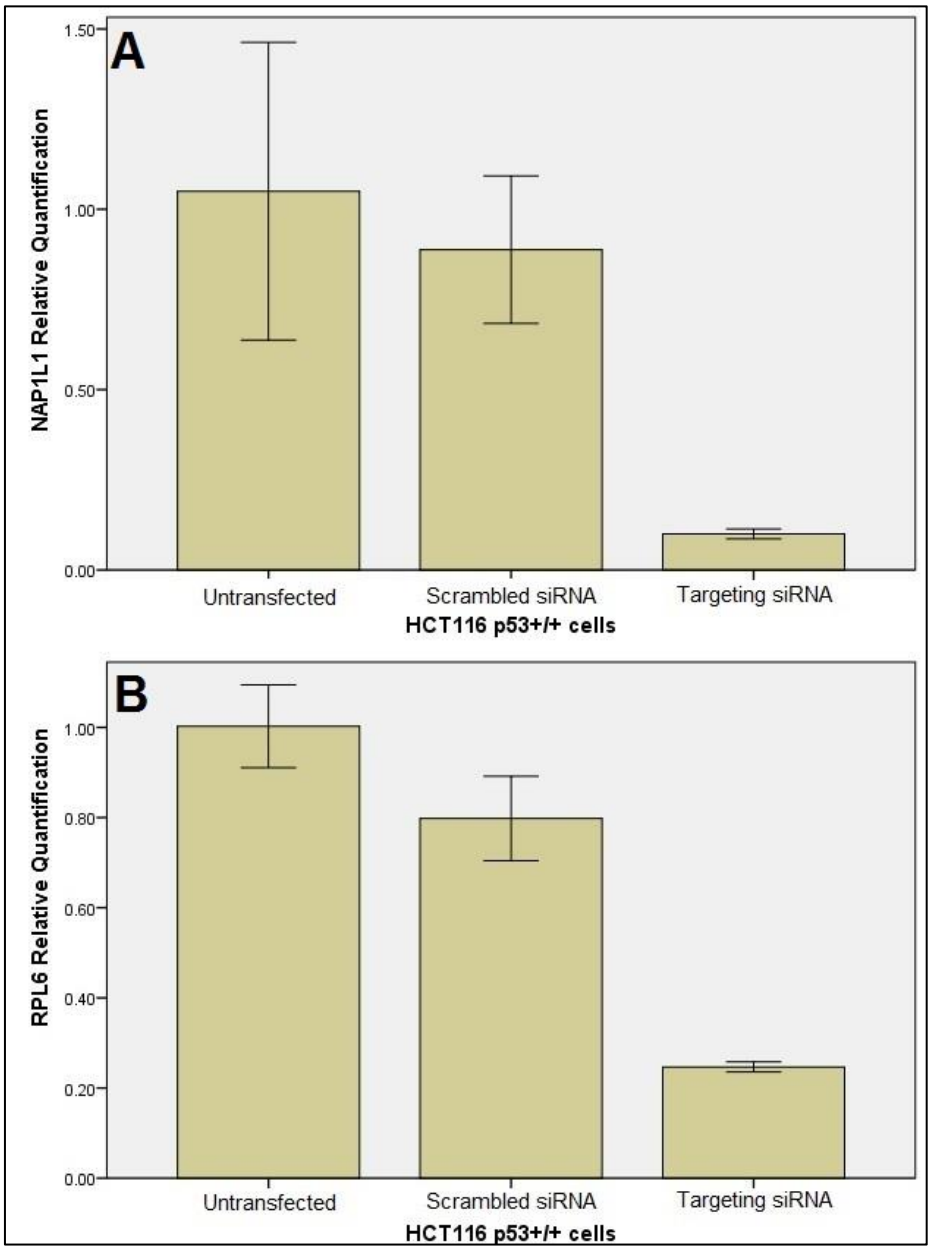


Figure 4.8. Assessment of gene expression interference efficiency in *TP53* wild-type HCT116 cells. *NAP1L1* mRNA levels (A) exhibited a 90% decrease in cells transfected with targeting siRNA whilst scrambled-siRNA-transfected cells showed only a 15% reduction. As for *RPL6* (B), transfection with targeting and non-targeting siRNA resulted in 75% and 20% gene silencing, respectively (error bars: +/-2SE; biological replicates: 2, technical replicates: 3).

Scrambled siRNA caused only a minor (less than 20%) decrease in gene expression when compared to non-transfected cells. These results confirm that successful and specific gene knockdown was achieved in *TP53* wild-type HCT116 cells.

As for *TP53* null HCT116 cells (figure 4.9), again a strong reduction in gene expression was demonstrated by the use of siRNAs targeting either *NAP1L1* or *RPL6*. However, in the case of *NAP1L1*, a 30% reduction in transcript levels was also noticed when using scrambled non-targeting siRNA, thus highlighting a moderate degree of non-specific silencing. This suggests that selective targeting might have occurred, making this scrambled siRNA a poor negative control for this experiment. Limited time and resources unfortunately prevented us from testing different non-targeting siRNAs in this study. *RPL6* transcript levels did not exhibit any reduction with scrambled siRNA in *TP53* null HCT116 cells. As specific silencing (caused by targeting siRNAs) was, in general, much more prominent than non-specific knockdown, we decided to proceed with the assessment of *TP53* null cells in the planned proliferation experiments.

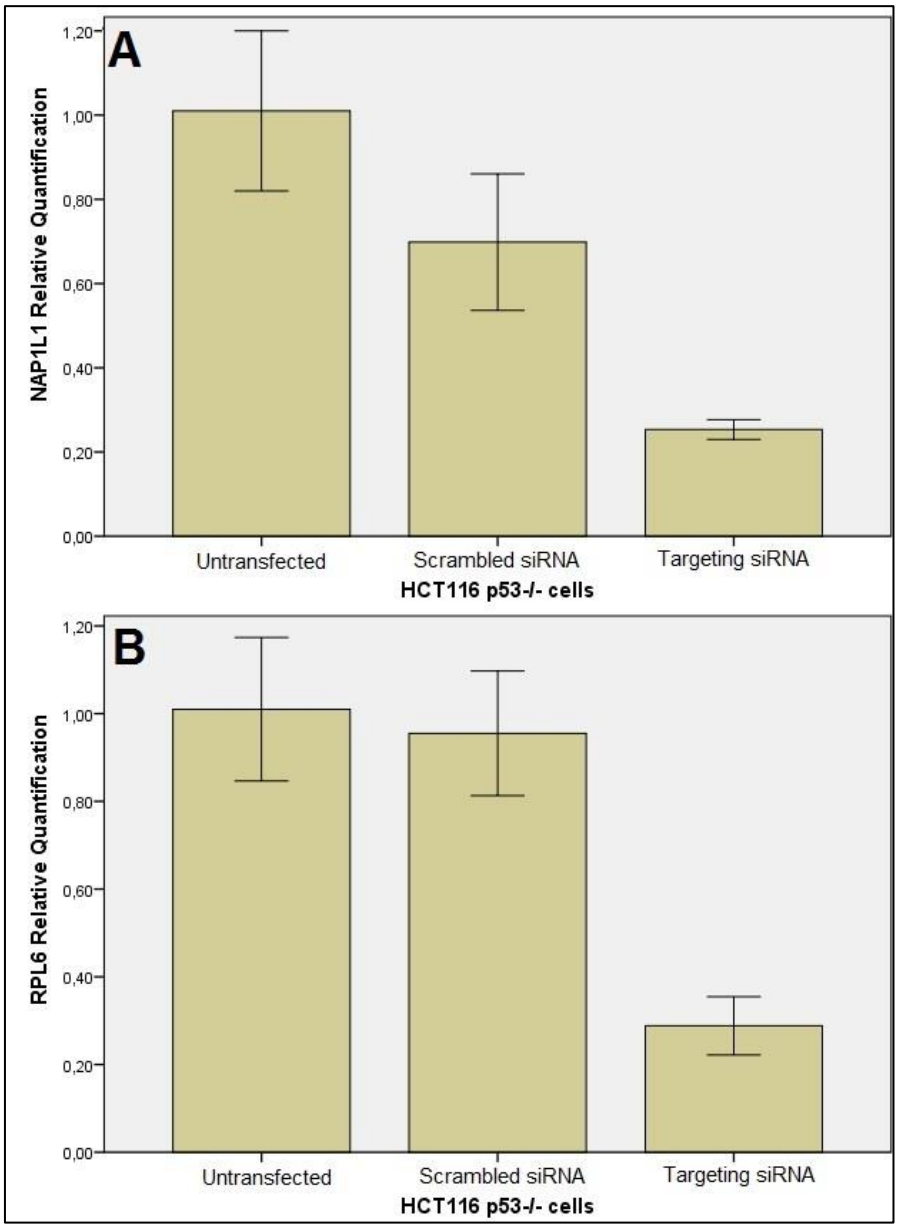


Figure 4.9. Assessment of gene expression interference efficiency in *TP53* null HCT116 cells. (A) *NAP1L1* exhibited more non-specific silencing as scrambled-siRNA produced a 30% mRNA level reduction. However, targeting siRNA showed a higher knockdown efficiency - 75% reduction. (B) *RPL6* exhibited a much more specific pattern of silencing. Transfection with targeting and non-targeting siRNA resulted in 71% and 4% gene silencing, respectively (error bars: +/-2SE; biological replicates: 2, technical replicates: 3).

4.6.2. **Effect of *NAP1L1* and *RPL6* silencing on cell proliferation**

Sulforhodamine B (SRB) assay is a widely used proliferation test based on the estimation of cellular density via the determination of total protein content in culture plate wells (Vichai and Kirtikara, 2006). SRB, a bright pink aminoxanthene dye with two sulfonic groups, binds to basic aminoacid residues under mild acidic conditions and dissociation occurs in basic conditions in a stoichiometric fashion (Skehan *et al*., 1990). Therefore, the amount of dye extracted from stained cells is proportional to the cell mass. Following the protocol detailed in *Chapter 2*, we performed SRB assays using HCT116 cells starting 48 hours after cell transfection.

***TP53* wild-type HCT116 cells**

In the SRB assay experiments, non-transfected and scrambled-siRNA-transfected *TP53* wild-type HCT116 cells showed a high proliferation rate, as denoted by the fast increase in cell density during the experiments (results in figure 4.10). *NAP1L1* siRNA transfection caused no effect on cell proliferation. Therefore, at least using the SRB technique, a role for *NAP1L1* in cell proliferation could not be confirmed. Conversely, *RPL6* siRNA provoked a strong and significant inhibition in cell growth from day 1 to 3. Non-targeting siRNA did not cause any degree of proliferation blockade, thus suggesting that *RPL6* silencing was the only factor responsible for this effect. This confirms the involvement of the *RPL6* gene in the proliferation of this CRC cell line.

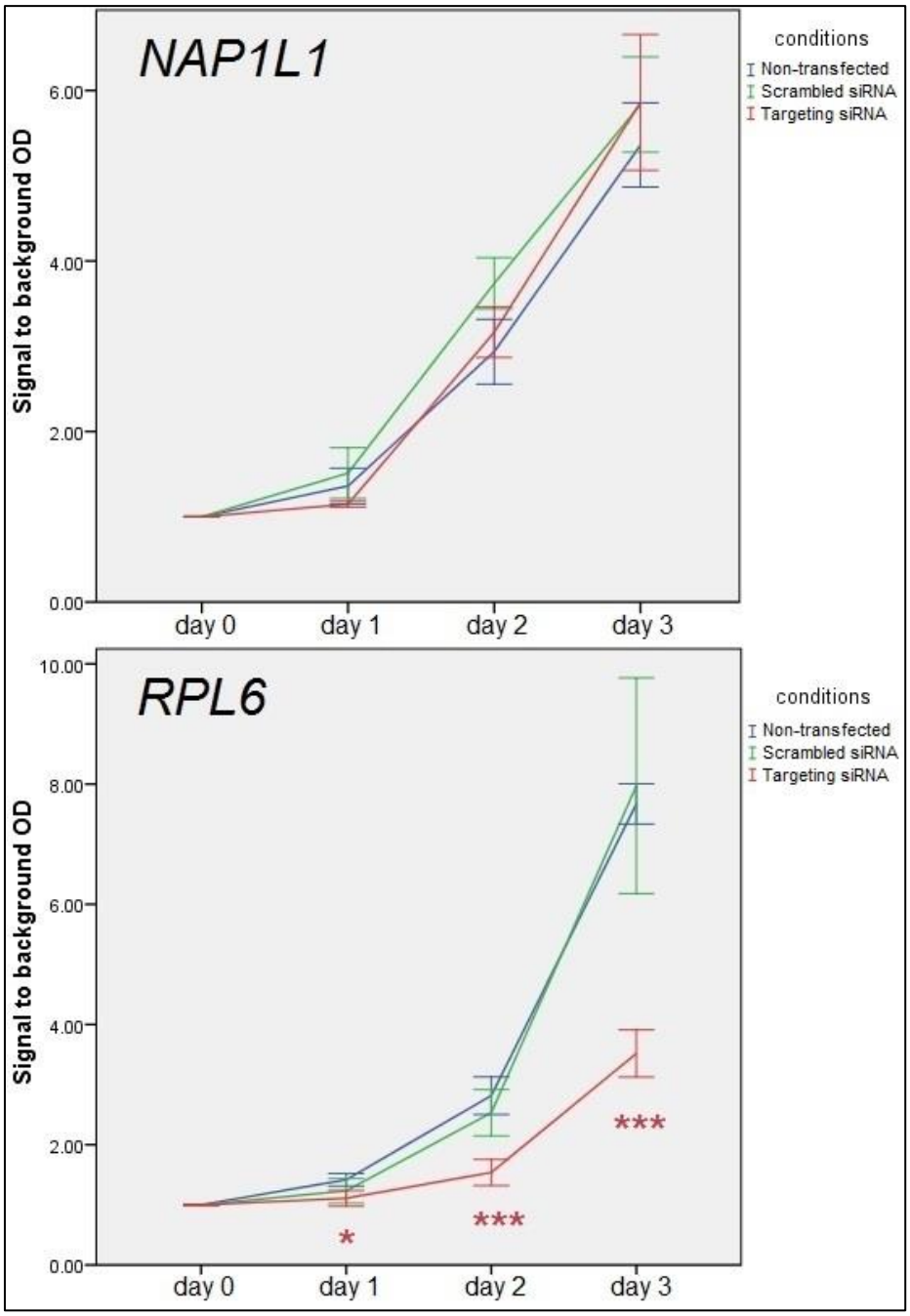


Figure 4.10. Effect of *NAP1L1* and *RPL6* knockdown on the proliferation of *TP53* wild-type HCT116 cells (SRB assay). *NAP1L1* silencing had no effect on cell proliferation, whilst *RPL6* knockdown resulted in clear inhibition of cell growth when compared to non-transfected cells or cells transfected with scrambled-siRNA (*p<0.05; ***p<0.001; ANOVA with post-hoc Dunnett T3 test; error bars: +/-2SE; N=1, n=3).

TP53 null HCT116 cells

The interpretation of the proliferation results in *TP53* null cells was more problematic. On superficial inspection, siRNA targeting either *NAP1L1* or *RPL6* resulted in significant inhibition of cell proliferation on day 2 of the SRB assay (figure 4.11). However, a careful analysis revealed some issues. First, the proliferation rates were considerably lower than those observed with *TP53* wild-type cells. No cell group exhibited more than a 2-fold increase in cell mass during the entire experiment. Additionally, cell density peaked at day 2, and then, decreased in the non-transfected and scrambled-siRNA transfected groups, whilst keeping low cell mass in the silenced group. The reason for this inhibition may be related to an increased sensitivity to nutrient deprivation in cells lacking a functional p53, given that the culture medium is not replaced during the proliferation assay. It has previously been demonstrated that *TP53* mutation or silencing result in increased cell death under glucose deprivation in melanoma (Chavez-Perez *et al*., 2014) and breast cancer cell lines (Rodriguez *et al*., 2012).

Together, these results permit us to conclude that *RPL6* gene silencing caused an important inhibition on the proliferation of *TP53* wild-type HCT116 cells. Conversely, *NAP1L1* knockdown had no effect upon the growth of the same cell line. No definite conclusion can be drawn regarding the effect of *NAP1L1* or *RPL6* knockdown on *TP53* null cells due to the general low rate of proliferation and increased amounts of cell death observed during these experiments. Therefore, only *TP53* wild-type HCT116 cells were used in the next part of this research – the assessment of the effect of *RPL6* silencing on the expression of known CRC-related genes.

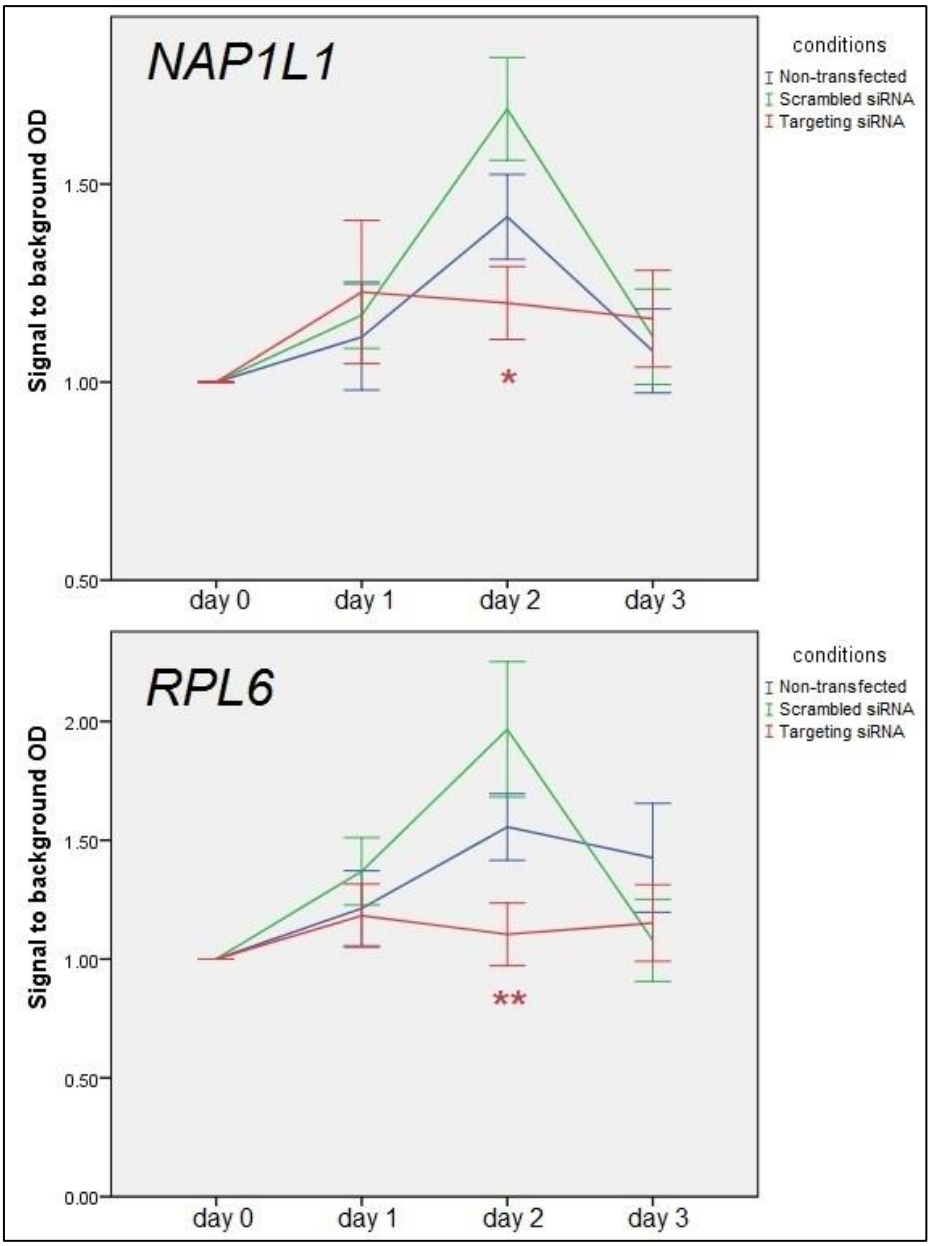


Figure 4.11. Effect of *NAP1L1* and *RPL6* knockdown on the proliferation of *TP53* null HCT116 cells (SRB assay). Minimal inhibition of cell growth was observed in cells transfected with siRNAs targeting both *NAP1L1* and *RPL6* on day 2 only. Noteworthy, the overall proliferation rate was exceptionally low for all conditions, as shown by the low increase in optical densities from days 0 to 3 (zero to 2.0-fold increase). Additionally, most conditions exhibited a reduction in cell mass by day 3. (* $p<0.05$; ** $p<0.01$; ANOVA with post-hoc Dunnett T3 test; error bars: +/-2SE; N=1, n=3).

4.6.3. **Effect of *RPL6* silencing on CRC-related genes**

RPL6 gene silencing caused a clear and strong inhibition of the proliferation of *TP53* wild-type HCT116 cells. Therefore, we hypothesised that this gene could influence the expression of other CRC-related genes and pathways. To assess this possibility, we extracted RNA from cells that had been transfected with scrambled and targeting siRNA. After the synthesis of cDNA, samples were loaded into Human Developmental Phases of Colorectal Cancer® array plates (Applied Biosystems, Carlsbad, CA, USA) along with the appropriate PCR master mix. This array plate assesses the expression of 28 CRC-related genes in addition to 4 endogenous controls (listed in table 4.2).

Table 4.2. Genes assessed in the CRC array plate.

18S	eukaryotic 18S rRNA – control
GAPDH	glyceraldehyde-3-phosphate dehydrogenase – control
HPRT1	hypoxanthine phosphoribosyltransferase 1 – control
GUSB	glucuronidase, beta – control
APC	adenomatous polyposis coli
BAX	BCL2-associated X protein
CTNNB1	catenin (cadherin-associated protein), beta 1
DCC	deleted in colorectal carcinoma
E2F4	E2F transcription factor 4
KRAS	Kirsten rat sarcoma viral oncogene homolog
MMP1	matrix metalloprotease 1
MMP2	matrix metalloprotease 2
MMP3	matrix metalloprotease 3
MMP7	matrix metalloprotease 7
MMP8	matrix metalloprotease 8
MMP9	matrix metalloprotease 9
MMP10	matrix metalloprotease 10
MMP11	matrix metalloprotease 11
MMP12	matrix metalloprotease 12
MMP13	matrix metalloprotease 13
MMP19	matrix metalloprotease 19
MMP20	matrix metalloprotease 20
MMP26	matrix metalloprotease 26
MMP28	matrix metalloprotease 28
MLH1	mutL homolog 1
MSH2	mutS homolog 2
MSH3	mutS homolog 3
MSH6	mutS homolog 6
PTGS2	prostaglandin-endoperoxide synthase 2
SMAD4	SMAD family member 4
TGFBR2	transforming growth factor, beta receptor II
TP53	tumour protein p53

Five target genes showed no amplification in the experiments (*DCC* and *MMPs 3, 8, 20* and *26*). The remaining 23 targets and 4 endogenous controls exhibited good amplification curves and were therefore included in the final analysis. Several methods for comparing relative quantification results are available, most of them involving fold-change, p-values produced by t-tests, or combinations of both (McCarthy and Smyth, 2009). As this was an exploratory analysis based on a single biological replicate (plated in triplicate), no or minimal within-group variability was expected. Therefore, simple fold-change was selected as the method for screening potentially altered genes. A two-fold increase or decrease in gene expression was considered the biologically relevant cut-off.

Relative quantification was calculated using the comparative CT method. Figure 4.12 depicts the relative expression of 23 CRC-related genes observed in *RPL6*-silenced *TP53* wild-type HCT116 cells compared to cells that had been transfected with scrambled-siRNA. Two genes exhibited increased expression above the selected fold-change cut-off: *BAX* (RQ: 2.3, fold-change: 2.3) and *MSH2* (RQ: 68, fold-change: 68). Conversely, two other genes showed relevant decreases in their relative expression: *MMP-12* (RQ: 0.0025, fold-change: - 400) and *MMP-13* (RQ: 0.28, fold-change:- 3.57). Some other genes also demonstrated minor variations in gene expression but these have not been included in this discussion.

As *RPL6* knockdown provoked an inhibition of cell proliferation in HCT116 cells, it is intuitive to suppose that this gene is involved in the malignant phenotype of these CRC cells. Although it might be the result of some direct anti-proliferative effect, a reduction in *RPL6* expression could cause an alteration in the expression of other genes or pathways associated with CRC growth. *BAX* and *MSH2* were shown to be up-regulated upon *RPL6* knockdown. The products of both genes are important players in physiological mechanisms of cancer control. BAX is a protein belonging to the BCL-2 family, a group of proteins that regulate apoptosis, specifically having pro-apoptotic properties (Czabotar *et al*., 2014). Apoptosis is one of the main regulators of cell proliferation and is an essential barrier against cancer development and progression. Supporting our finding, RPL6 has been

previously shown to inhibit BAX expression in gastric cancer cells (Du *et al*., 2005). The importance of BAX as an anti-cancer factor is highlighted by the attempts to induce the up-regulation of this protein as a form of cancer therapy (Cory and Adams, 2005, Liu *et al*., 2016). The role of MSH2 in the protection against carcinogenesis has already been discussed in *Chapter 1*. It is part of the DNA mismatch repair (MMR) system, a mechanism of defence against errors in DNA replication frequently inactivated in colorectal cancers (Sameer *et al*., 2014). *MSH2* expression was strikingly up-regulated in *RPL6*-silenced cells (68-fold). This suggests that the over-expression of *RPL6* observed in colorectal tumours and the adjacent non-neoplastic mucosa (shown earlier in this chapter) might result in a strong inhibition in *MSH2* expression, thus favouring the occurrence of potentially carcinogenic mutations. To the best of our knowledge, there has been no previous report of an association of *RPL6* and either the apoptosis pathway or the DNA-mismatch repair system in CRC.

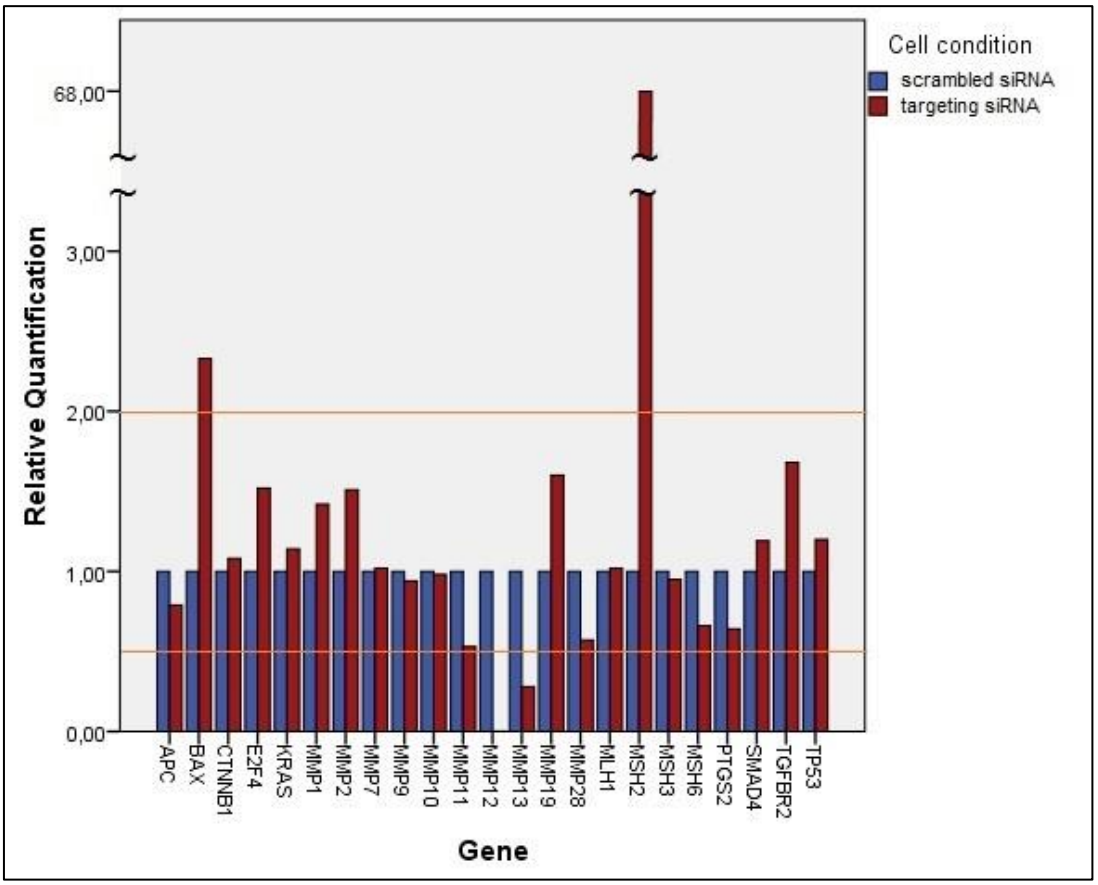


Figure 4.12. The effect of *RPL6* gene silencing on the expression of CRC-related genes in *TP53* wild-type HCT116 cells. Horizontal lines represent the fold-change cut-offs (+/- 2-fold-change). *BAX* and *MSH2* exhibited increased gene expression, whilst *MMP-12* and *MMP-13* showed decreased transcriptional levels upon *RPL6* silencing (N=1, n=3).

Matrix metalloproteinases (MMPs) are proteins that are associated with neoplastic matrix remodelling, cellular invasion and angiogenesis (Gialeli *et al*., 2011), and have recently been recognised as having several other functions from cancer initiation to metastasis (Shay *et al*., 2015). The involvement of MMPs in CRC has been extensively studied and several reviews about the subject are available (Herszenyi *et al*., 2012, Zucker and Vacirca, 2004). Our study has shown that the expressions of *MMP-12* and *MMP-13* were reduced after *RPL6* knockdown (see figure 4.12 above). Increased *MMP-12* expression has been associated with adverse prognostic factors and reduced survival in cancers of the lung (Hofmann *et al*., 2005), stomach (Zheng *et al*., 2013) and melanomas (Zhang *et al*., 2015b). Moreover, *MMP-12* polymorphism has been linked to an increased risk of

advanced stage at diagnosis in patients with CRC (VAN Nguyen *et al*., 2013). Similarly, *MMP-13* expression has also been linked to worse prognosis and metastasis in various cancers such as gastric (del Casar Lizcano *et al*., 2003), oral (Huang *et al*., 2016) and breast (Zhang *et al*., 2008) carcinomas, and sarcomas (Zyada and Shamaa, 2008). Regarding CRC, *MMP-13* over-expression has consistently been associated with advanced tumour stage, metastasis and poor survival (Yang *et al*., 2012, Yamada *et al*., 2010, Leeman *et al*., 2002). Collectively, these findings point to an important role of these two MMPs during colorectal carcinogenesis. Thus, the resultant decrease in the expression of *MMP-12* and *MMP-13* upon *RPL6* knockdown would result in an inhibitory effect on cellular features important to the malignant phenotype. No association between *RPL6* and metalloproteinases has previously been reported.

4.7. Discussion

Quantitative PCR has become a popular method for translational research in oncology. It can be used for screening of potentially useful biomarkers and for validation of candidates identified by proteomic or genomic approaches (Skrzypski, 2008, Clark-Langone *et al*., 2007, Scott *et al*., 2011). An example of the successful use of qPCR in this setting is the development of the *Oncotype Dx*, a qPCR-based panel of breast cancer biomarkers capable of defining the risk of recurrence after surgical treatment of early stage tumours (Paik *et al*., 2004). Based on the recurrence score provided by the test, clinicians can tailor treatments and avoid unnecessary toxic therapies in low-risk patients. This test has been widely used for node-negative, hormone receptor-positive breast cancer. The company responsible for the development of *Oncotype Dx* has also created a similar test for colon cancer, the *Oncotype Dx Colon Cancer Assay* (Clark-Langone *et al*., 2007, Clark-Langone *et al.*, 2010). However, this version of the test has not gained as much acceptance as the breast cancer assay, mainly due to the lack of robust clinical evidence of benefit from its use (Webber *et al*.,

2010). Therefore, the discovery of CRC biomarkers by qPCR is still an open and promising research field.

The selection of the candidate biomarkers under study here was based mainly on the results from previous studies in our group performed by Abeer Hammoudi, who showed increased expression of *RPL6* and *PHB* (Hammoudi *et al*., 2013) in animal models of intestinal carcinogenesis, and by Shahram Ibrahim, who suggested *NAP1L1* as a potential CRC biomarker (Ibrahim, 2014). *CTNNB1* was also tested, in order to better clarify the Wnt pathway status in our samples. The qPCR experiments were performed after testing the conditions in several optimisation and quality-control steps previously described.

The expression of *CTNNB1* was similar in normal controls and adjacent tissues, and was only slightly increased in tumour samples. As mentioned earlier, the few reports describing *CTNNB1* expression in CRC have shown contradictory results (Anwar *et al*., 2015, Qin *et al*., 2006, Truant *et al*., 2008). Gene mutations, rather than increased gene expression, appears to be relevant in colorectal carcinogenesis, as several reports have consistently demonstrated the presence of *CTNNB1* mutations in some cases of sporadic CRC (Cancer Genome Atlas Network, 2012, Anwar *et al*., 2015). Moreover, activation of the Wnt signalling pathway in CRC is more commonly a result of *APC* mutations resulting in accumulation of β-catenin via decreased degradation than a consequence of increased mRNA expression.

Regarding our candidate biomarkers, we observed a strikingly similar pattern of expression for *NAP1L1* and *RPL6*, a correlation that was also observed in the immunohistochemical work presented in *Chapter 3*. Both genes were shown to be highly expressed in both the tumour and the adjacent mucosa from patients with CRC when compared to normal controls. These findings suggest that these genes may be involved in “field cancerisation” or “field effect”, a phenomenon first introduced by Slaughter *et al*. in 1953 (Slaughter *et al*., 1953) and well described in cancer development. It is largely attributed to genetic and epigenetic modifications leading to

altered gene expression (Baba *et al*., 2016), making tissues adjacent to tumours prone to develop new primary cancers. In one report assessing CRC tissues, epigenetic modifications were found in adjacent tissues collected either 2cm or 8cm from the primary tumour (Park *et al*., 2016). We analysed adjacent tissues collected 10cm or more from the tumour and still observed a differential expression suggesting that environmental factors might affect large portions of the intestine. Although we have not specifically assessed epigenetic modifications, gene expression is directly or indirectly affected by such alterations, suggesting this mechanism as a possible explanation for the results observed. *NAP1L1* has been shown to have important functions in maintaining dedifferentiation in cancer-derived (Li *et al*., 2012) and pluripotent stem cells (Gong *et al*., 2014). It also regulates proliferation in murine stem cells (Yan *et al*., 2016). Taken together, these results suggest that *NAP1L1* might be a fundamental player in cancer development and progression. *RPL6* gene expression and function have been far less explored in cancer research. Studies using gastric cancer cell lines have suggested an important role in proliferation and multi-drug resistance (Du *et al*., 2005, Wu *et al*., 2011, Gou *et al*., 2010). Additionally, this protein has been shown to interact with the MDM2-p53 complex upon ribosomal stress, operating as a feedback loop for p53 activity in embryonic and cancer cells lines (Bai *et al*., 2014). The paucity of data regarding *RPL6* function in carcinogenesis impairs the formulation of a precise hypothesis to explain our findings. P53 loss-of-function might explain the increased expression of *RPL6* in cancer tissues (based on the feedback loop just explained). However, it would not provide an explanation for the increased expression in the adjacent tissues, as *TP53* mutations are considered late events in colorectal carcinogenesis.

Differences in demographic characteristics between groups should also be taken into account when interpreting these findings. In this study, groups were well balanced in terms of age. Thus, genetic and epigenetic changes due to the aging process are unlikely to have influenced the results. All samples were collected in the same region, thus minimising the possibility of ethnicity misbalance. On the other hand, there was a predominance of men in the cancer cohort whilst women formed the majority of the normal

control individuals. Although no relationships between gender, sex hormones and the expression of our candidates have been reported, further research involving gender-matched cohorts could clarify this issue. There are other factors related to the procedures used for sample collection (i.e. colonoscopy versus surgery) that might also explain the observed differences in gene expression between normal and cancer/adjacent samples. For example, the bowel preparation and type of anaesthesia were different between these two procedures. The potential impact of these differences in our results would also require additional research.

Another point of discussion is the apparent discordance between the findings of the immunohistochemical and the gene expression studies that we report. Both NAP1L1 and RPL6 exhibited a reduction in nuclear immunostaining in cancer tissues when compared to the adjacent unaffected colonic mucosa, whilst these genes showed increased expression in both tumour and adjacent samples, compared to normal controls. Concordance between mRNA expression and protein content has long been demonstrated to be poor as a result of post-transcriptional, post-translational and protein degradation regulation (Vogel and Marcotte, 2012). A recent large-scale analysis of 16,561 genes and corresponding proteins in 200 normal and cancer samples showed a *Spearman´*s Rho correlation coefficient of only 0.45 (Kosti *et al*., 2016). Only 6.1% of the gene/protein pairs exhibited statistically significant correlations (Rho values ranging from 0.77 to 1.0). None of the genes tested in this section of our research were listed in this group, thus providing support for the RNA/protein dissociation that we observed. Besides the mechanisms mentioned above, extracellular translocation may also explain the absence of protein accumulation within the cells. The results from Line *et al*. showing that NAP1L1 is an antigenic protein that produces an immune response in a subset of patients with CRC but not in healthy individuals provide the rationale for this hypothesis (Line *et al*., 2002). The dissection of the precise mechanisms underlying NAP1L1 and RPL6 cellular distribution is beyond the scope of this research. Nonetheless, the assessment of the expression of these proteins in blood derived from CRC patients and healthy individuals might result in a clinically useful

biomarker with potential implications for cancer screening, diagnosis and surveillance. Therefore, the assessment of NAP1L1 and RPL6 protein concentration in blood using immuno-based assays was undertaken and the details of these experiments are described in the next chapter.

Based on the differential expression of *NAP1L1* and *RPL6* between tumour/adjacent tissues and normal controls, we decided to perform a study of the mechanistic consequences of gene silencing on cell proliferation and on the expression of other CRC-related genes. By using siRNAs, we successfully blocked the transcription of these two genes in a colorectal cancer cell line (*TP53* wild-type and null HCT116 cells). Next, we carried out experiments to assess the effect of gene silencing on cell proliferation. *NAP1L1* silencing did not produce any alteration of cell proliferation in *TP53* competent cells. The literature suggests that *NAP1L1* regulates the differentiation and proliferation of stem cells (Li *et al*., 2012, Gong *et al*., 2014, Yan *et al*., 2016) and pancreatic neuroendocrine tumours (Schimmack *et al*., 2014). Therefore, altered *NAP1L1* expression in CRC might not be associated with proliferation but with the maintenance of an undifferentiated status. Differently, *RPL6* silencing resulted in strong inhibition of cell growth in the same cell type. A similar effect has already been demonstrated in gastric cancer cell lines (Wu *et al*., 2011, Du *et al*., 2005), reinforcing the importance of this gene for the proliferation of cancer cells. HCT116 cells with deleted *TP53* demonstrated a more erratic pattern of proliferation when transfected with scrambled or targeting siRNAs. Moreover, a low basal proliferation rate (even in the non-transfected group) and possible cell toxicity were noticed during the proliferation assay. These issues prevented us from drawing any definite conclusion regarding the silencing of *NAP1L1* and *RPL6* in cells without a functioning *TP53*.

RPL6 knockdown of *TP53* wild-type cells was further analysed in order to assess the effects of gene silencing on the expression of CRC-related genes. Using qPCR-based techniques, we demonstrated that, among 23 genes known to be associated with CRC, *RPL6* knockdown resulted in increased expression of *BAX* and *MSH2* and decreased expression of *MMP-12* and *MMP-13*. *BAX* and *MSH2* are important players in protecting against

cancer development, whilst MMPs are effectors in cancer invasion, metastasis and angiogenesis, among other features associated with a malignant phenotype. Taken together, these results suggest some possible mechanisms by which *RPL6* up or down-regulation could control CRC cell behaviour, thus explaining the influence of this gene on cell proliferation and, possibly, other malignant features. Although these findings are hypothesis generating, further research is strongly recommended in order to clarify the role of *RPL6* during carcinogenesis in general and in CRC in particular.

PHB is a ubiquitous protein with an essential role in protection against oxidative stress, a factor known to be involved in carcinogenesis (Jones, 2008). Other cellular functions regulated by PHB include proliferation and survival via interactions with signalling cascades such as the Ras-Raf-MEK-Erk pathway and caspases (Chowdhury *et al*., 2014). In our study, *PHB* gene expression results did not suggest a clear role of this gene in CRC. Although a trend pointing to increased transcript levels in adjacent and tumour tissues was observed, this was not statistically significant. The small sample size might explain this negative result. However, the results from Hammoudi *et al*. showed only a 1.32-fold increase in *PHB* expression in tumour versus adjacent samples (Hammoudi *et al*., 2013). This may suggest that *PHB* mRNA over-expression is not as marked in CRC as it has been demonstrated in other cancer types (Franzoni *et al*., 2009, Jiang *et al*., 2013, Kang *et al*., 2008, Ummanni *et al*., 2008). Consequently, we decided not to further study PHB gene function using additional methods in this project. However, the differential cellular distribution of this protein revealed by our immunohistochemical studies deserves further clarification. We will therefore explore possible associations between PHB immuno-expression and clinical and prognostic variables in a dedicated chapter (*Chapter 6*), along with the other promising markers: NAP1L1 and RPL6.

Chapter Five:

Assessing blood circulating candidate protein biomarkers for CRC: enzyme-linked immuno-assay analysis (ELISA) and development of an electrochemiluminescence (ECL) assay for NAP1L1

5. CHAPTER 5 – ASSESSING BLOOD CIRCULATING CANDIDATE PROTEIN BIOMARKERS FOR COLORECTAL CANCER: ENZYME-LINKED IMMUNO-ASSAY ANALYSIS (ELISA) AND DEVELOPMENT OF AN ELECTROCHEMILUMINESCENCE (ECL) ASSAY FOR NAP1L1

5.1. Introduction

The results described up to this point have clearly shown a differential expression of our candidate protein biomarkers in tissues from CRC patients. However, the immunohistochemistry and qPCR experiments both required tissue fragments for analysis. These samples, in turn, require invasive procedures such as endoscopy or surgery. Ideally, a biomarker test should provide diagnostic, prognostic or predictive information in the least harmful manner. In this regard, blood tests are favoured as being minimally-invasive and are a widely accepted way of obtaining biological samples. For example, a German study has shown that only 37% of patients offered screening colonoscopy accepted this test. From the population that refused colonoscopy, only 15% accepted an alternative stool test, whilst 83% accepted a blood test for *SEPT9* (Adler *et al*., 2014). In fact, blood collection is part of the routine workup of most health conditions. The uncomplicated collection and storage of blood derivatives such as plasma and serum together with the comprehensive proteome represented in these biological fluids ensure that it will remain the preferred diagnostic material for the foreseeable future (Rifai *et al*., 2006).

In the biomarker discovery workflow (see figure 5.1), an initial extensive prospection of candidate biomarkers using methods such as liquid chromatography tandem mass spectrometry (LC-MS/MS) is usually performed (Rifai *et al*., 2006). This process is known as an “unbiased” approach, as there are no anticipated biomarker candidates. Subsequently, the proteins that exhibit differential expression are further assessed using more targeted techniques. This is the pathway that we have been following in this research after the initial findings published by Hammoudi (Hammoudi *et al*., 2013). The first method that was devised to accurately measure the

concentrations of peptides and proteins in biological fluids was radioimmunoassay (RIA). It was first reported in 1960 (Yalow and Berson, 1960) and revolutionised laboratory medicine by allowing the precise quantification of important substances. Nonetheless, the use of radioactively labelled substances requires complex facilities to decrease occupational exposure and environmental contamination. To address this issue, Perlmann and Engvall in Sweden, and Schuurs and van Weemen in the Netherlands conceptualised and developed modifications of the RIA and replaced the radioactive compound with an enzyme as the reporter label (Engvall and Perlmann, 1971, Van Weemen and Schuurs, 1971). In the late 1970s and early 1980s, enzyme immunoassays (EIA) finally matched the RIA´s exquisite sensitivity. This achievement, coupled with the practicality and the possibility of automation soon made ELISAs and other non-radioactive EIAs the most used methods for protein quantification in biological fluids (Lequin, 2005). Currently, ELISAs are one of the most widely employed approaches for verification and validation of new biomarkers (Del Campo *et al*., 2015, Rifai *et al*., 2006).

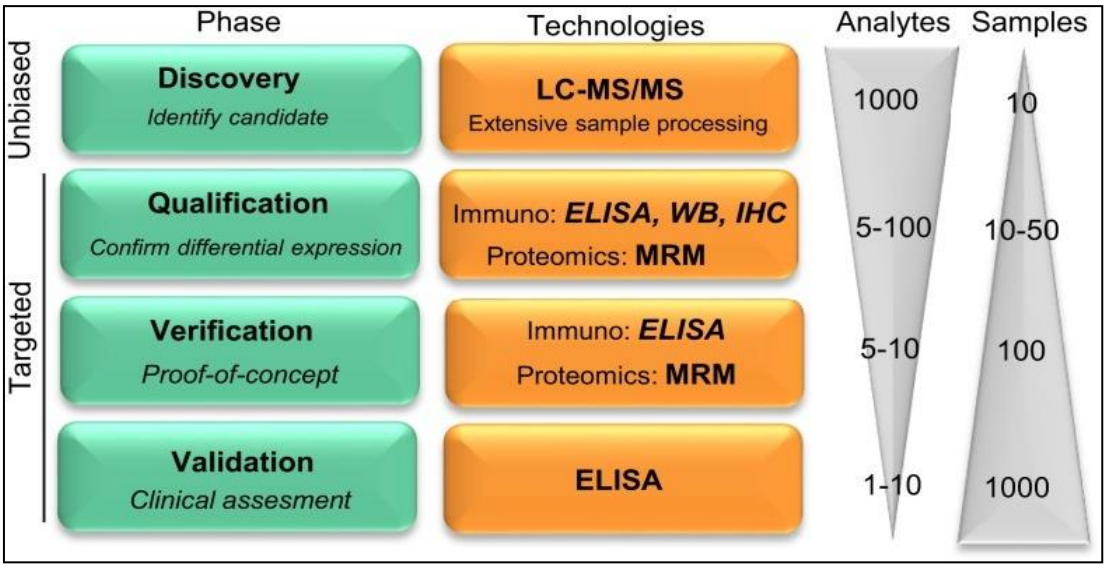


Figure 5.1. Main phases of biomarker discovery and validation. The process starts with a low-throughput technique (i.e. using few samples) that is capable of analysing hundreds or thousands of analytes, such as LC-MS/MS. This is called the "unbiased" step. Then, more targeted methods are used in order to confirm the differential expression of the candidate biomarkers using larger numbers of samples. LC-MS/MS, liquid chromatography tandem mass spectrometry; MRM, multiple reaction monitoring; IHC, immunohistochemistry; WB, Western blotting. Figure from (Del Campo *et al*., 2015), adapted from (Rifai *et al*., 2006).

There are a number of different ELISA formats in laboratory practice. The sandwich-ELISA is the most commonly used due to its high sensitivity and specificity (Price and Newman, 1997). In this format, two antibodies (a capture antibody and a detection antibody) are used. Although immunoassays are deemed ideal methods for biomarker validation in biological fluids, the performance of the test is highly dependent upon several steps during the development and optimisation of the assay. A wide range of variables from antibody specificity to the composition and concentration of the different reagents can affect the final result (Del Campo *et al*., 2015). If the required sensitivity is not achieved with the enzymatic method, alternative reactions such as fluorescence or chemiluminescence may be used (Price and Newman, 1997). Additionally, the best matrix for immunoassays, plasma or serum, is highly variable and depends on each specific analyte (O'Neal *et al*., 2014, Alsaif *et al*., 2012). Therefore, the appropriate matrix should be determined for each protein prior to any validation study.

Several CRC-associated biomarkers have been assessed in blood. Fung *et al*. analysed 32 proteins as potential biomarkers and found that three of them, Insulin like growth factor binding protein 2 (IGFBP2), Dickkopf-3 (DKK3), and Pyruvate kinase M2 (PKM2), constituted a diagnostic panel for early-stage CRC with a performance superior to FOBT (Fung *et al*., 2015). Pengjun *et al*. demonstrated that a panel of cytokines, CEA and CA 19-9 measured by ELISA could assist in CRC detection (Pengjun *et al*., 2013). Several authors have also shown that circulating methylated *Septin 9* can also be used for CRC diagnosis (deVos *et al*., 2009, Grutzmann *et al*., 2008, Toth *et al*., 2012). However, none of the suggested panels have been thoroughly assessed in large validation cohorts and the real benefit of their use remains unproven. Here, we decided to test the proteins that had been identified in previous work from our group. The concentrations of RPL6 and PHB have previously been demonstrated to be increased in blood from animal models of intestinal carcinogenesis (Hammoudi *et al*., 2013). NAP1L1, a protein differentially expressed in tissues from both animal models of CRC and human cancer samples (Ibrahim, 2014), was also shown to be increased in a preliminary study from our group (data discussed in the next section). Therefore, we undertook an ELISA study of these proteins to investigate whether they were differentially expressed in the blood of CRC patients. This could potentially lead to the development of a new panel of CRC biomarkers.

Commercial ELISA kits have theoretically been tested and optimised by their manufacturers. However, laboratory conditions may vary and the performance of these kits should be externally validated in order to guarantee their accuracy. A good way to assess the quality of a kit is by checking the reproducibility of the results from previous studies which have used the test in different settings. Nonetheless, this is not always possible as many proteins are poorly studied and, sometimes, have not been assessed at all in published papers. When validated kits are not available, the assessment of new candidate biomarkers requires the development and optimisation of novel kits. Commercial kits for NAP1L1, RPL6 and PHB are currently provided by some manufacturers. However, none of our candidate

biomarkers has previously been well studied using ELISA. Searching the English-language literature using *PubMed*, *Scopus* and *Web of Science* databases with the keywords "*ELISA*", "*enzyme-linked immunosorbent assay*", "*immunoassay*" and the protein names, we found no report of NAP1L1, RPL6 or PHB assessment using ELISA in cancer or in any other health condition. Only a single Chinese-language study had evaluated PHB in cervical carcinoma patients using ELISA (only the abstract is available in English). The authors reported a decreased concentration of PHB in the serum of cervical cancer patients compared to normal controls (Yan *et al*., 2013). Therefore, no standard commercial kit was identified for any of our candidate biomarkers and no literature background was found to guide our experiments. Consequently, we decided to start our assessment using kits that had previously been tested by our group in some preliminary experiments (described below). Subsequently, we also decided to test both biofluids – serum and plasma, in order to improve the yield of valid results.

5.2. Preliminary ELISA data

The initial experiments assessing ELISA kits to measure the candidate biomarkers were performed by Nadeem Al-Khafaji as part of his MRes project in our research group (unpublished work). In this evaluation, he prospectively collected serum samples from individuals who underwent colonoscopies at the Countess of Chester Hospital (Chester, UK) and at the Royal Liverpool University Hospital "Inflammatory Bowel Disease Clinic" (Liverpool, UK). As this analysis aimed mainly to evaluate the suitability of the ELISA kits, few colorectal cancer cases were included. Individuals providing serum for this assessment had the following endoscopy/pathology diagnoses: normal colon (12 cases), Crohn´s disease (10), ulcerative colitis (11), low-grade dysplastic colonic adenomas (4), current CRC (1) and previously resected CRC with a suspected recurrence (1). For the purpose of hypothesis generation, the two later cases were combined together (cancer group) and were compared to the remaining non-malignant groups.

The kits used for this analysis were designed by Cloud-Clone Corp (Houston, USA) and assembled by Uscn Life Science Inc (Wuhan, China). These providers were chosen because their kits had previously been tested by Hammoudi *et al*. using blood from animal models of CRC with promising results (Hammoudi *et al*., 2013) and because they manufactured kits for all the candidate proteins that we needed to test: NAP1L1, RPL6 and PHB. After initial optimisations, serum samples were diluted to 1:3 in sterile PBS and were tested in duplicate using the protocol recommended by the manufacturer and described in *Chapter 2*. Plate reading was carried out using a Tecan Sunrise™ 96-well plate reader at 450nm.

The initial purpose of this experiment was to determine whether or not protein concentrations would lie within the detection range of the kit. For this sample set, all kits were able to detect the serum concentration of all samples tested. Next, the concentrations of the proteins in each group were compared to provide insights regarding possible trends and guide the following stages of our investigation. When the concentration of the proteins was collectively plotted as a panel and the individual groups were compared, we observed a trend towards overall increased expression in the CRC group compared to the non-malignant groups (see figure 5.2A). The comparison between malignant and non-malignant groups for individual proteins exhibited the same trends for NAP1L1 and PHB (figure 5.2B and C) but not for RPL6 (figure 5.2D).

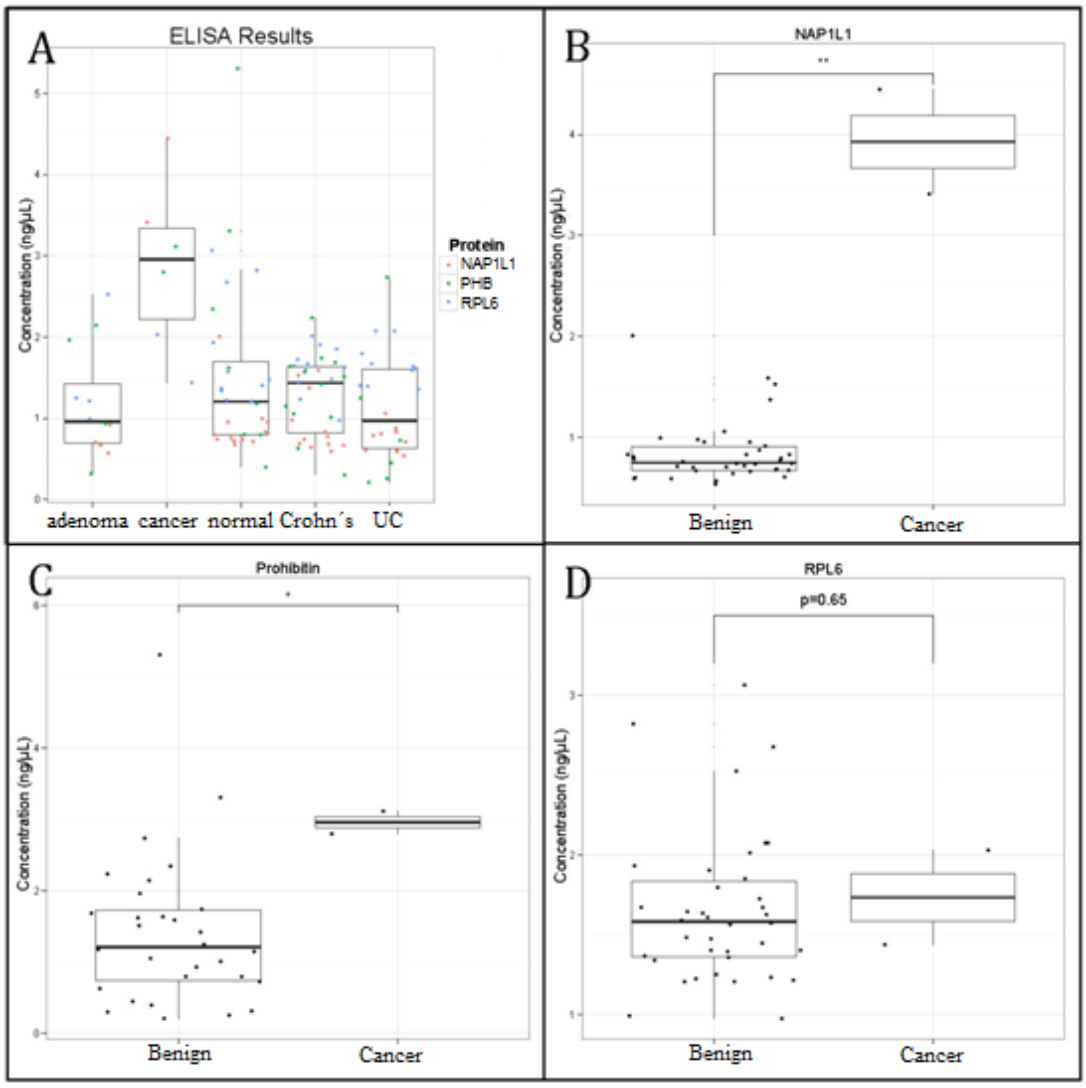


Figure 5.2. Preliminary ELISA results for the assessment of NAP1L1, PHB and RPL6 in serum samples from individuals with CRC (2 cases) and several non-malignant conditions (37 cases). When plotted as a panel, a trend towards increased expression of the proteins was observed (A). Although the limited number of samples prevents any solid conclusion at this stage, the comparison between the malignant and non-malignant conditions showed increased NAP1L1 and PHB expressions in CRC cases (B and C). No such trend was observed for RPL6 (D). *p<0.05; **p<0.01 (Mann-Whitney U test). Reproduced with permission from Nadeem Al-Khafaji.

These initial findings provided the basis for continuing our assessment of these candidates in a validation phase. Based on these results, we decided to test the same ELISA kits using another sample cohort which included more CRC and colonic adenoma cases. We also decided to omit samples from patients with inflammatory bowel diseases, as no trend had

been observed in this group of samples. Therefore, normal individuals, individuals with colorectal adenomas and CRC patients were included.

5.3. Clinical samples

For this investigation, we used a new set of serum and plasma samples collected at the Countess of Chester Hospital from normal individuals, individuals with adenomatous polyps and CRC patients – the UK cohort. Samples were also collected from similar groups of patients in the city of Cuiaba, Brazil, forming the Brazilian cohort. UK samples were collected during the years 2015/2016 whilst Brazilian samples were collected from 2013 until mid-2015. Similar protocols were used for sample collection, processing and storage of the two cohorts. Noteworthy, Brazilian samples were shipped to the UK under optimal conditions (dry ice) using a company specialised in the transportation of biological samples. Thus, the main difference between these two groups was the length of storage, although minor differences during collection and processing cannot be excluded as local technicians prepared most of the samples. Some samples were used during the initial tests and assay optimisations. Given the various procedures involved in these phases, several samples were therefore totally used and, consequently, were not available for the final experiments. Therefore, we kept a set of samples from each cohort unused until the final experiments, so that we could produce final results for the same samples in each part of this study. In order to reassure that only clinical groups with a representative number of samples were assessed, we decided to test only normal controls and cancer samples for the Brazilian cohort, as there were few adenoma cases within this sample set. For the UK cohort, given the presence of larger numbers of samples in each group, we compared normal controls, low- and high-grade adenomas and cancer samples. Table 5.1 describes the demographic characteristics of the individuals from the Brazilian cohort who provided the samples that were tested in the final immunoassay experiments carried out in this study. Unfortunately, the same information was not available for the individuals from the UK cohort due to the fact that this

investigator is not registered as a clinician in the UK and did not have access to clinical information for UK patients.

Table 5.1. Clinical and demographic characteristics of the individuals from the Brazilian cohort included in the final immunoassay experiments. Mean age for cancer group was significantly higher than for normal individuals.

Characteristics	**Normal control (n=10)**	**Cancer (n=30)**	**p**
Mean age (range)	43 (23-56)	55 (34-83)	0.009*
Gender			NS
Male	4	16	
Female	6	14	
Stage			NA
I-II	-	47%	
III-IV	-	53%	

* t- test. NS: non-significant. NA: not applicable.

5.4. ELISA testing using commercial kits

In this section, we describe the issues and results that were obtained when commercial ELISA kits were used to test these sample cohorts. As mentioned earlier, our starting point was the platform from Cloud-Clone Corp tested preliminarily. Although we used similar protocols and conditions as before, we encountered several difficulties when trying to reproduce the results that have previously been demonstrated. Among the most serious issues, low sensitivity (many samples lying below the detection range) and poor inter-plate consistency (similar kits yielding different results for the same samples) were the most relevant. In order to overcome these problems, we tested kits purchased from different manufacturers and also a different sample matrix (plasma). In the following sub-sections, a summary of these experiments and the results for each protein are provided. As the comprehension of the issues and findings observed requires a detailed

understanding of individual ELISA results for particular cases, tables displaying results for all samples tested in each relevant experiment are provided. When appropriate, graphs showing mean concentrations and statistical information are also shown.

5.4.1. **NAP1L1 kits – testing and results**

The Cloud-Clone Corp NAP1L1 ELISA kit (product number SEH571Hu) served as our starting point in this assessment. NAP1L1 isoform-1 (Homo sapiens) is made of 391 amino acids. According to the manufacturer (personal communication), the immunogenic sequence for this kit ranges from Ala2 to Lys197. The detection range for this kit is reported to be 0.119 – 20.0ng/mL. Using the protocols suggested by the manufacturer and the sample conditions used in the preliminary phase described earlier, we attempted to replicate the initial results. Testing ten serum samples from each of the four groups (normal, low-grade adenoma, high-grade adenoma and cancer) in duplicate, we obtained the results depicted in table 5.2. As can be seen, different from the preliminary test results, various samples exhibited optical readings below the detection range of the kit, thus preventing any comparison between groups. In order to rule out any experimental error, a similar kit from the same provider was tested with other samples from this same cohort. Although meticulous care was taken during sample preparation and assay procedures, similar results were obtained (these results will be shown later when the comparison between serum and plasma is discussed). Once the possibility of technical error had been excluded, another issue that might have played a part was sample collection and storage. However, in this experiment we used samples collected both in Chester-UK (labelled with the initials JJ) and in Cuiaba-Brazil (all the remaining samples). Brazilian samples were collected between 1 and 3 years prior to the experiments being undertaken; Chester samples were collected only weeks or a few months prior to the analysis, and both were stored at -80ºC until use. The initial results for both groups were very similar, thus

suggesting that inadequate sample preparation and storage were unlikely sources of inconsistency. Based on these results, we concluded that this kit did not exhibit the necessary consistency and dynamic range for the validation of our candidate biomarkers. Therefore, different kits needed to be tested.

Table 5.2. Initial ELISA results for NAP1L1 using the Cloud-Clone kit.

Group	Sample Id	Concentration (ng/mL)
Normal	N2	2.332
	N3	0.506
	N4	0.368
	N5	6.678
	N6	**invalid value**
	N7	0.956
	N8	0.248
	N11	0.458
	N22	**invalid value**
	N23	0.49
Low-grade adenoma	P2	0.568
	P3	1.098
	P4	**invalid value**
	P5	0.768
	P6	0.276
	P7	1.316
	P8	**invalid value**
	P9	**invalid value**
	P10	0.854
	P18	0.99
High-grade adenoma	JJ047	**invalid value**
	JJ058	**invalid value**
	JJ093	0.154
	JJ111	**invalid value**
	JJ138	1.22
	JJ156	**invalid value**
	JJ165	**invalid value**
	CR68	0.732
	CR78	**invalid value**
	CR95	1.14
Cancer	CR62	0.466
	CR65	**invalid value**
	CR67	0.278
	CR70	**invalid value**
	CR73	0.274
	CR76	1.934
	CR80	0.676
	CR83	0.422
	CR85	**invalid value**
	CR86	**invalid value**

Ten serum samples from each of the four groups were tested. A large proportion of the samples exhibited readings below the detection range of the kit (assigned as “invalid value”). Thus, no average concentration could be calculated and no comparison between groups could be carried out. This performance was different from the results obtained using the same kit in the preliminary phase.

In order to clarify whether the poor results were due to the suboptimal performance of that specific kit, we decided to test plates from other manufacturers that used different protein epitopes as the immunogen. This represented quite a challenge as all of the contacted providers reported that their kits used the same or largely overlapping areas of the NAP1L1 N-terminal region. One of the only kits that targeted a different area of the protein was provided by DL Develop – product number DL-NAP1L1-Hu (Wuxi, China). The immunogen used in this kit spans from *Ala2* to *Thr289,* thus encompassing a larger area of NAP1L1 than the Cloud-Clone kit described above. Its product sheet describes performance parameters (such as detection range, sensitivity and precision) very similar to those provided by the Cloud-Clone kit. Assay procedures recommended by the manufacturers were also identical.

Using the same samples tested in the previous experiment, we assessed the performance of the DL Develop NAP1L1 ELISA kit. This step aimed mainly to investigate whether or not this kit had a better sensitivity than the one provided by Cloud-Clone Corp, thus being able to define NAP1L1 concentrations for all samples. The results for this test are shown in table 5.3. Differently from the Cloud-Clone kit, the DL Develop NAP1L1 kit produced optical density readings within the detection range for all cases and concentrations were defined for all serum samples. However, a careful analysis of these results revealed some additional problems. Firstly, most of the concentrations were low, close to the lower limit of detection for the kit. This finding was different from the results produced in the preliminary phase, although it must be highlighted that a different kit was then used. Additionally, despite observing a trend of increased NAP1L1 concentration in the serum of CRC patients, this was not statistically significant (see figure 5.3).

Table 5.3. ELISA results for NAP1L1 using the DL Develop kit.

Group	Sample Id	Concentration (ng/mL)	Median (ng/mL)
Normal	N2	0.186	
	N3	0.172	
	N4	0.17	
	N5	0.297	
	N6	0.387	
	N7	0.266	
	N8	0.206	
	N11	0.216	
	N22	0.194	
	N23	0.194	**0.2288**
Low-grade adenoma	P2	0.218	
	P3	0.309	
	P4	0.198	
	P5	0.156	
	P6	0.284	
	P7	0.21	
	P8	0.316	
	P9	0.288	
	P10	0.194	
	P18	0.21	**0.2383**
High-grade adenoma	JJ047	0.234	
	JJ058	0.204	
	JJ093	1.06	
	JJ111	0.19	
	JJ138	0.176	
	JJ156	0.258	
	JJ165	0.23	
	CR68	0.267	
	CR78	0.168	
	CR95	0.198	**0.2985**
Cancer	CR62	0.238	
	CR65	0.198	
	CR67	0.236	
	CR70	0.706	
	CR73	1.024	
	CR76	0.132	
	CR80	0.26	
	CR83	0.262	
	CR85	0.346	
	CR86	0.321	**0.3723**

The results are different from the Cloud-Clone kit, even though these are the same samples. All the samples produced readings within the detection range of this assay. Notably, all calculated concentrations were very low compared with the preliminary results. Median concentrations for each group were calculated and compared, as seen in the figure below.

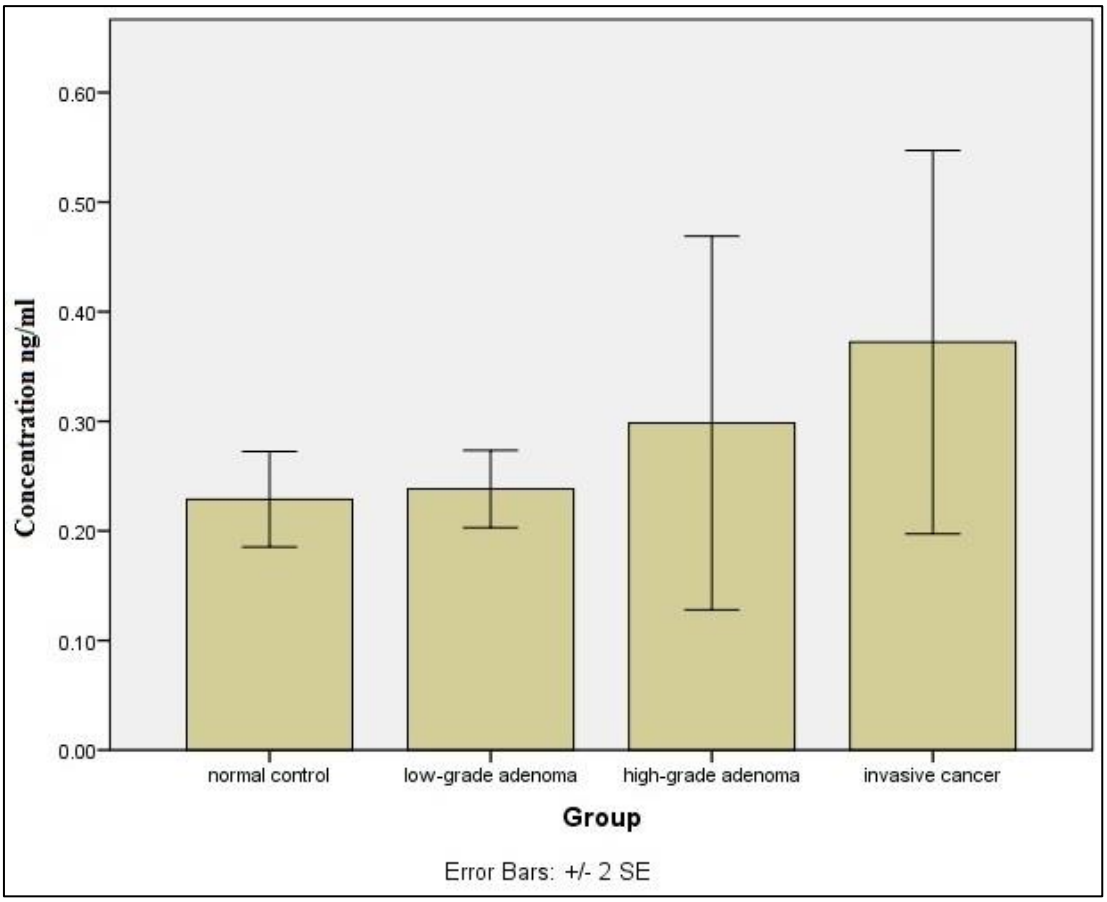


Figure 5.3. Graphic representation of NAP1L1 serum concentrations using the DL Develop kit. A trend of increased concentration was observed in serum samples from high-grade adenomas and CRCs. However, this difference was not statistically significant (p=0.287, Kruskall-Wallis test). N=10 cases per group.

Next, we directly compared the results produced by the two kits that we tested, as we used the same samples and the same assay procedures in both experiments. If concordant, we could conclude that the accuracies of both kits were similar, and therefore, we could opt for the DL Develop plate due to its better detection range and lower limit of detection. Unfortunately, however, this was not the case. As table 5.4 depicts, the concordance between the two plates was extremely poor. Some samples exhibiting higher concentrations in the Cloud-Clone plate showed low concentrations in the DL Develop plate, and vice versa. Calculated *Pearson´s* correlation coefficient was r = - 0.13, denoting an insignificant negative correlation between the results produced by the two assays.

Table 5.4. Comparison of results obtained with NAP1L1 ELISA plates from Cloud-Clone and from DL Develop. Some samples were retested in another Cloud-Clone plate to assess the consistency of that kit.

Sample	1st Cloud-Clone plate (ng/mL)	2nd Cloud-Clone plate (ng/mL)	DL Develop plate (ng/mL)
N2	2.332	1.86	0.186
N3	0.506	-	0.172
N4	0.368	**invalid value**	0.17
N5	6.678	-	0.297
N6	**invalid value**	-	0.387
N7	0.956	-	0.266
N8	0.248	-	0.206
N11	0.458	-	0.216
N22	**invalid value**	-	0.194
N23	0.49	-	0.194
P2	0.568	-	0.218
P3	1.098	-	0.309
P4	**invalid value**	-	0.198
P5	0.768	-	0.156
P6	0.276	-	0.284
P7	1.316	-	0.21
P8	**invalid value**	-	0.316
P9	**invalid value**	-	0.288
P10	0.854	-	0.194
P18	0.99	-	0.21
JJ047	**invalid value**	-	0.234
JJ058	**invalid value**	**invalid value**	0.204
JJ093	0.154	0.346	1.06
JJ111	**invalid value**	**invalid value**	0.19
JJ138	1.22	1.28	0.176
JJ156	**invalid value**	**invalid value**	0.258
JJ165	**invalid value**	-	0.23
CR68	0.732	-	0.267
CR78	**invalid value**	-	0.168
CR95	1.14	-	0.198
CR62	0.466	-	0.238
CR65	**invalid value**	-	0.198
CR67	0.278	-	0.236
CR70	**invalid value**	-	0.706
CR73	0.274	-	1.024
CR76	1.934	-	0.132
CR80	0.676	-	0.26
CR83	0.422	-	0.262
CR85	**invalid value**	-	0.346
CR86	**invalid value**	-	0.321

As also shown in the table above, seven serum samples were re tested in duplicate in another Cloud-Clone kit in order to ascertain the inter-assay coefficient of variation (reported by both manufacturers as being <12%). Coefficient of variation is computed by the formula: CV (%) = (SD/mean) x 100, where SD is the standard deviation (Connett and Lee, 1990). Four samples showed concentrations below the detection range in one or both plates. The three samples with valid results exhibited a CV of 25.3% between plates. Therefore, although higher than the CV reported by the manufacturers, the consistency between the two Clone plates was better than the concordance between Cloud-Clone and DL Develop plates.

The aforementioned results did not allow us to confidently select a platform to proceed. At this point, therefore, we decided to test a different biological fluid (plasma), to check whether this would produce more consistent results. For this experiment, we selected 18 cases representing normal controls, high-grade adenomas and CRC. Samples were chosen based on the availability of both serum and plasma in sufficient quantities to be used in this and in subsequent tests, if necessary. Serum and plasma from each case were tested in duplicate using the Cloud-Clone NAP1L1 ELISA kit. This experiment also intended to test the CV of this kit, as described above, because we re-tested 7 samples. Strikingly, this analysis revealed that all plasma samples exhibited valid readings, whilst the serum samples repeated the pattern of a high proportion of readings being below the detection range (see table 5.5). Furthermore, whenever valid readings were obtained for both plasma and serum samples for a particular case, the former was higher than the latter. This remarkable finding imposed a totally new perspective for our investigation. It suggested that plasma instead of serum should be used in any further experiment for the assessment of NAP1L1. This finding, if confirmed, may also have implications for future studies assessing the concentration of this protein in blood samples.

Table 5.5. Comparison of NAP1L1 results between plasma and serum samples using a Cloud-Clone ELISA kit. Whilst serum samples performed similarly to previous observations using the same plate, all plasma samples yielded valid results. Additionally, whenever concentrations for both plasma and serum were valid, plasma concentrations were higher than serum concentrations.

	Plasma		**Serum**	
Groups	**Samples**	**(ng/mL)**	**Samples**	**(ng/mL)**
Normal Chester	JJ001P	0.902	JJ001S	**invalid value**
	JJ002P	1.544	JJ002S	0.438
	JJ007P	0.396	JJ007S	**invalid value**
Normal Brazil	N2P	2.698	N2S	1.86
	N4P	0.536	N4S	**invalid value**
High-grade Chester	JJ058P	0.71	JJ058S	**invalid value**
	JJ093P	1.682	JJ093S	0.346
	JJ111P	0.712	JJ111S	**invalid value**
	JJ138P	2.338	JJ138S	1.28
	JJ156P	0.744	JJ156S	**invalid value**
Cancer Chester	JJ057P	1.692	JJ057S	**invalid value**
	JJ113P	0.78	JJ113S	**invalid value**
	JJ126P	0.842	JJ126S	**invalid value**
	JJ143P	1.2	JJ143S	**invalid value**
	JJ145P	0.862	JJ145S	**invalid value**
	JJ166P	1.564	JJ166S	**invalid value**
	JJ079P	0.732	JJ079S	**invalid value**
	JJ121P	1.508	JJ121S	**invalid value**

Although the demonstration of higher concentrations of NAP1L1 in plasma samples certainly opens an exciting new scientific avenue, the comparison of the median concentrations between patient groups did not show any noticeable difference, at least in this very small sample set (figure 5.4). Any conclusion could however only be made after testing a larger sample set and using a consistent assay. For this purpose, we chose to proceed with testing plasma samples only. As none of the kits that we tested (Cloud-Clone and DL Develop) proved to be ideal platforms, we decided to test a third kit. This time, a UK-based provider with substantial expertise in manufacturing immunoassays – Abbexa Ltd, was chosen.

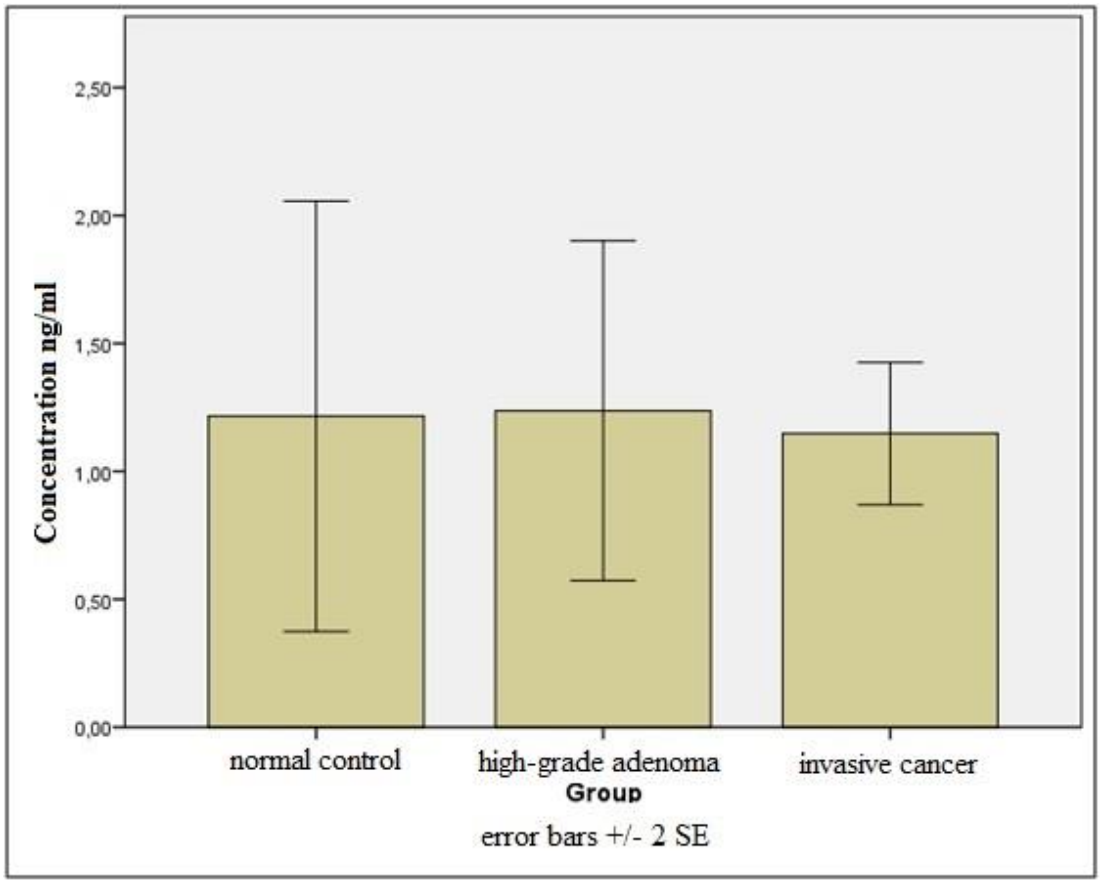


Figure 5.4. Mean plasma NAP1L1 concentration results using the Cloud-Clone ELISA kit. The comparison of the results between the groups did not show a significant difference (p=0.866, Kruskal-Wallis test). However, the limited number of samples per group prevented any robust conclusion. N= 5 normal controls, 5 high-grade adenomas and 8 invasive cancer cases.

The NAP1L1 ELISA kit from Abbexa – product number abx152455 (Cambridge, UK) targets the protein area between *Ala2* to *Lys197* (personal communication), the same area used as the epitope for the Cloud-Clone kit. Assay procedures and performance features reported by the manufacturer were essentially the same as those reported for the previous two kits. Using 96-well plates, we aimed to test plasma samples from the most representative groups for each cohort (UK and Brazil). In order to further assess whether or not the length of storage had an effect on the results, we analysed each cohort separately, as the UK cohort had a significantly shorter sample storage time than the Brazilian cohort. Brazilian samples available for this assessment were mainly normal controls and CRC cases. Thus we decided to focus on assessment of those two groups. For the UK cohort, as a larger number of samples from patients who had adenomas were present,

we tested samples from all four groups (normal controls, low-grade adenomas, high-grade adenomas and CRC).

Plasma samples from the Brazilian cohort showed several results below the detection range of the kit, as shown in table 5.6. However, the proportion of invalid values was lower than that observed when we tested serum samples. Additionally, taking only valid results into account, the mean concentrations per group were considerably higher than those observed when testing serum samples. These findings reinforced the notion that NAP1L1 concentrations are easier to measure in plasma than in serum. However, comparison of the mean concentration of the valid results did not show any difference between normal controls and CRC cases (figure 5.5). A wide range of results was observed within each group, suggesting that NAP1L1 levels are variable and heterogeneous in both normal individuals and CRC patients. Noteworthy, the lowest valid result in this test was 0.822ng/mL, a value well above the claimed inferior limit of detection of the kit (0.105ng/mL, according to the manufacturer). It suggests that the performance of this kit, at least in our laboratory, was inferior to the specifications provided by the company.

Table 5.6. Abbexa NAP1L1 ELISA kit results using plasma samples from the Brazilian cohort.

NORMAL (n=10) Sample Id	Concentration (ng/mL)	CANCER (n=30) Sample Id	Concentration (ng/mL)
N1P	2.205	CR60P	2.997
N2P	**invalid value**	CR62P	3.078
N3P	0.897	CR63P	3.102
N4P	**invalid value**	CR64P	2.079
N5P	**invalid value**	CR65P	1.86
N6P	5.268	CR67P	1.155
N7P	11.904	CR68P	**invalid value**
N8P	1.872	CR69P	1.677
N11P	16.473	CR70P	5.133
N28P	16.482	CR71P	1.383
		CR72P	2.277
		CR73P	**invalid value**
		CR74P	0.822
		CR75P	1.548
		CR76P	**invalid value**
		CR77P	**invalid value**
		CR78P	1.776
		CR79P	1.35
		CR97P	1.452
		CR81P	**invalid value**
		CR82P	11.73
		CR84P	7.191
		CR85P	3.261
		CR86P	21.903
		CR87P	3.441
		CR88P	6.411
		CR89P	0.855
		CR90P	18.27
		CR91P	1.197
		CR92P	3.681

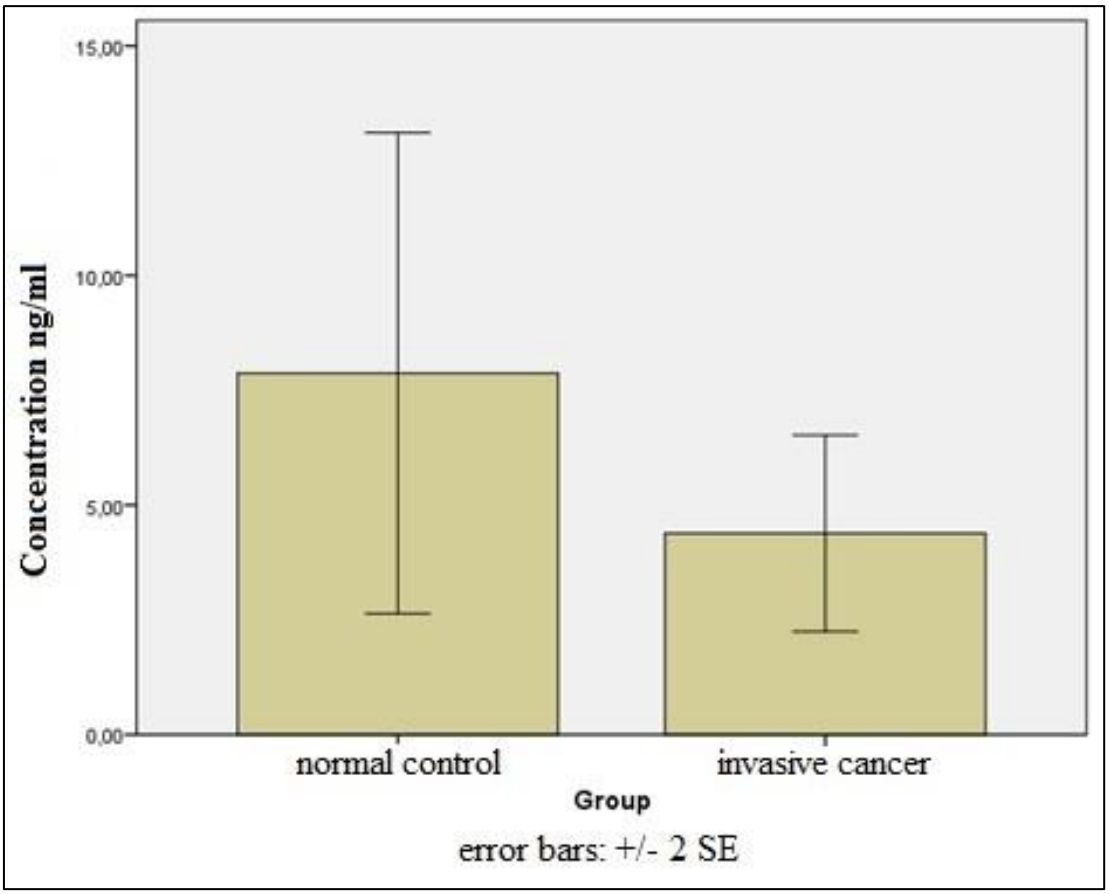


Figure 5.5. Mean NAP1L1 concentration in plasma samples from the Brazilian cohort (Abbexa kit). This graph was plotted excluding the invalid results in order to facilitate a visual comparison. No significant difference was demonstrated. Indeed, a trend of decreased NAP1L1 concentration in the cancer group was observed. (p=0.242, Mann-Whitney U test) N: 10 normal controls, 30 invasive cancers.

Our next task was to test the UK cohort using the same ELISA kit. We tested samples from 10 normal controls, 10 patients with low-grade adenomas, 7 patients with high-grade adenomas and 13 CRC patients. Results are shown in table 5.7 and in figure 5.6, below. From the 40 samples tested, only two exhibited readings below the detection range. This was less than the number of invalid samples observed in the Brazilian cohort (shown above). Additionally, the mean values in the UK-groups were higher than those seen in the Brazilian groups. In this experiment, the mean concentration was not statistically different between the four groups. However, a trend towards increased NAP1L1 concentration in the cancer group was noticed. An identical trend was previously observed when testing serum samples using the DL Develop kit (see figure 5.3), the only kit able to

produce valid results for all the serum samples. However, the magnitude of the readings was lower in that case, probably due to the use of serum instead of plasma.

Table 5.7. Abbexa NAP1L1 ELISA kit results using plasma samples from the UK cohort. Only two invalid results were produced. Concentrations were higher than those observed with the Brazilian cohort.

Normal (n=10)		**Low-grade (n=10)**		**High-grade (n=7)**		**Cancer (n=13)**	
Sample Id	**(ng/mL)**	**Sample Id**	**(ng/mL)**	**Sample Id**	**(ng/mL)**	**Sample Id**	**(ng/mL)**
JJ001	1.461	JJ005	2.268	JJ047	**Invalid**	JJ057	8.655
JJ002	14.058	JJ006	14.331	JJ058	13.221	JJ079	6.885
JJ007	6.216	JJ010	5.229	JJ093	19.311	JJ113	40.647
JJ012	11.496	JJ011	2.655	JJ111	14.745	JJ121	5.772
JJ013	14.469	JJ019	13.278	JJ138	7.527	JJ126	16.515
JJ015	23.724	JJ021	10.752	JJ156	14.037	JJ132	33.804
JJ016	0.666	JJ022	17.505	JJ165	3.681	JJ136	13.932
JJ018	1.575	JJ023	19.68			JJ143	12.795
JJ020	17.682	JJ041	**Invalid**			JJ144	12.894
JJ024	15.255	JJ050	8.103			JJ145	12.777
						JJ149	17.724
						JJ157	6.288
						JJ166	2.748

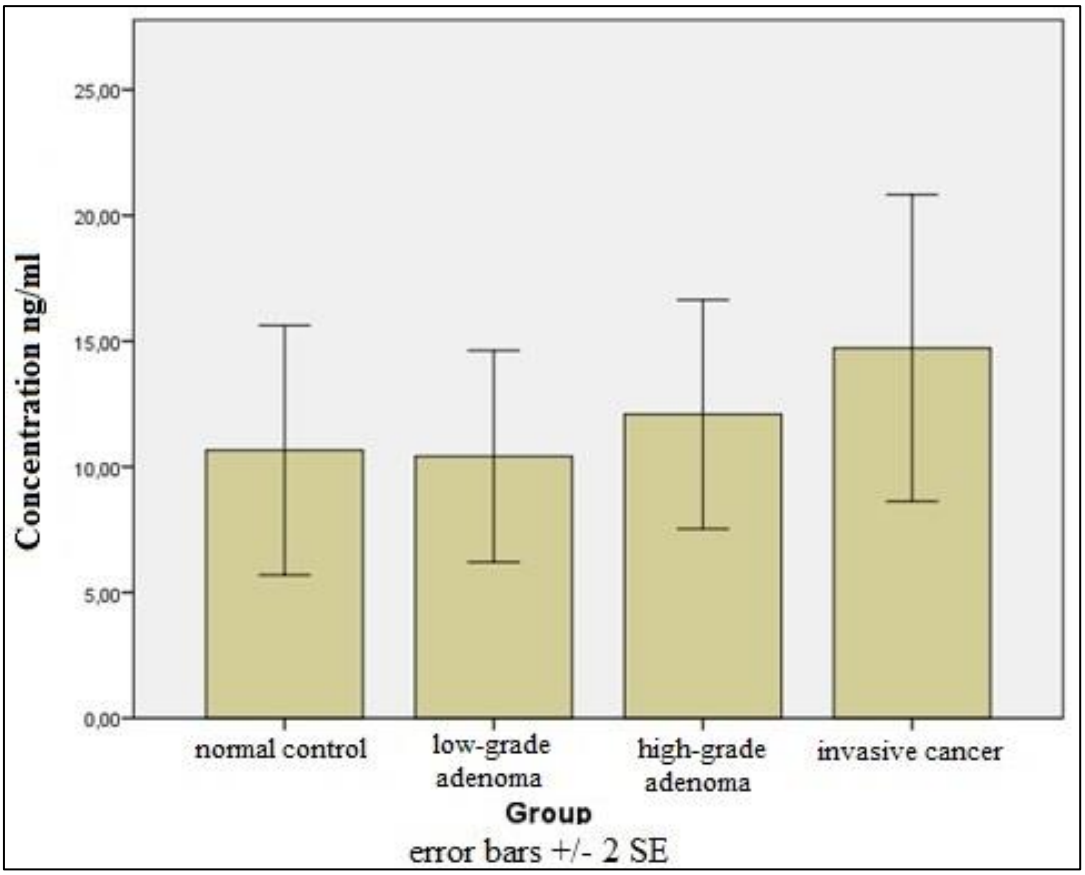


Figure 5.6. Mean NAP1L1 concentration in plasma samples from the UK cohort (Abbexa kit). A non-significant trend towards increased concentration in cancer samples was observed (p=0.885, Kruskal-Wallis test). A similar trend was seen when testing serum using the DL Develop kit (figure 5.3). Notably, these two experiments demonstrated the best sensitivities (fewer or no invalid values produced) among all the plates tested. N: 10 normal, 10 low-grade adenomas, 7 high-grade adenomas and 13 cancer samples.

Together, these results suggest that the length of storage may, indeed, affect NAP1L1 concentration in blood samples. Alternatively, some other unclear factor related to sample collection or processing might also play a part, as the Brazilian and the UK cohorts were obviously collected in different settings, although following similar protocols. This could even explain the discrepancy of these results when compared with our preliminary data. In those experiments, Al-Khafaji analysed serum that had been collected only a few weeks earlier, whilst here we used samples that had been kept in freezers for time periods ranging from several months to a few years. Although statistical significance was not observed, a trend towards increased NAP1L1 levels in cancer patients was observed in those

experiments which had better performances – the ones that produced fewer or no results below the lower limit of detection. The lack of adequate immunoassays and the use of samples with relatively long storage time could have prevented us from identifying a clear differential concentration of this protein in cancer patients. Any conclusions could only be confidently made with the use of accurate methods of protein measurement. Based on these assumptions, we realised that we had two possible ways to proceed with this investigation: we could either collect new samples prospectively and test them within a few weeks; or we could develop an assay which had better sensitivity and consistency than the ELISA kits that we had used until now. The first option was deemed unrealistic due to time constraints associated with this research. Additionally, we would still need to sort out issues related to the inconsistency of the kits. For these reasons, we chose the second option and set out to develop an in-house immunoassay for NAP1L1.

For this purpose, we selected the Meso Scale Discovery® (MSD) electrochemiluminescence (ECL) platform. Chemiluminescence methods are supposedly more sensitive than ELISAs (Price and Newman, 1997). MSD – ECL assays have been demonstrated to outperform ELISAs in the assessment of several proteins (Kuhle *et al*., 2016, Guglielmo-Viret and Thullier, 2007, Leary *et al*., 2013). Details of this method, the assay development process and the quality control tests are described in section 5.5, together with the results of the analysis of our samples.

Before discussing these ECL assays, we will describe the results of measuring RPL6 and Prohibitin concentrations in the blood using ELISA kits.

5.4.2. **RPL6 kits – testing and results**

Similarly to NAP1L1 evaluation, the starting point for RPL6 analysis was the ELISA kit produced by Cloud-Clone Corp that was used in the preliminary study (product number SEF046Hu). This kit uses antibodies targeting the area between *Ala2* and *Phe288* (personal communication), almost the entire protein which has 288 amino acids. According to the product manual, the lower limit of detection for this kit is 0.061ng/mL, considerably lower than the limit for the NAP1L1 kit from the same manufacturer (0.119ng/mL). Therefore, a better performance would be expected. Serum samples from the four groups were initially tested. Indeed, as can be seen in table 5.8, the lower limit of detection of this kit allowed valid results to be detected for most of the serum samples tested. Interestingly, all four invalid results were observed in the high-grade adenoma group (three samples from the UK and one sample from Brazil). The reason for this finding is not clear. The comparison between the remaining three groups showed similar RPL6 concentrations in normal samples, low-grade adenomas and cancer samples. In fact, a non-significant trend towards a decreased concentration was observed in the cancer group (see figure 5.7).

Table 5.8. Initial ELISA results for RPL6 using the Cloud-Clone kit.

Group	Sample Id	Concentration (ng/mL)	Median (ng/mL)
Normal	N2	3.123	
	N3	2.724	
	N4	1.552	
	N5	3.645	
	N6	1.827	
	N7	1.796	
	N8	2.286	
	N11	7.558	
	N22	1.24	
	N23	0.7	2.6451
Low-grade adenoma	P2	1.598	
	P3	9.312	
	P4	3.882	
	P5	1.386	
	P6	1.37	
	P7	2.838	
	P8	0.944	
	P9	1.71	
	P10	1.068	
	P18	1.816	2.5924
High-grade adenoma	JJ047	0.768	
	JJ058	0.378	
	JJ093	**Invalid**	
	JJ111	0.744	
	JJ138	1.364	
	JJ156	**Invalid**	
	JJ165	**Invalid**	
	CR68	4.089	
	CR78	1.75	
	CR95	**Invalid**	**Invalid**
Cancer	CR62	2.026	
	CR65	2.464	
	CR67	1.972	
	CR70	2.558	
	CR73	2.678	
	CR76	0.702	
	CR80	2.17	
	CR83	1.588	
	CR85	0.456	
	CR86	2.643	1.9257

Serum samples were tested and the four groups were compared. Four out of 40 samples demonstrated readings below the detection range, all of which were in the high-grade adenoma group. This performance was better than that demonstrated by the NAP1L1 kit from the same manufacturer.

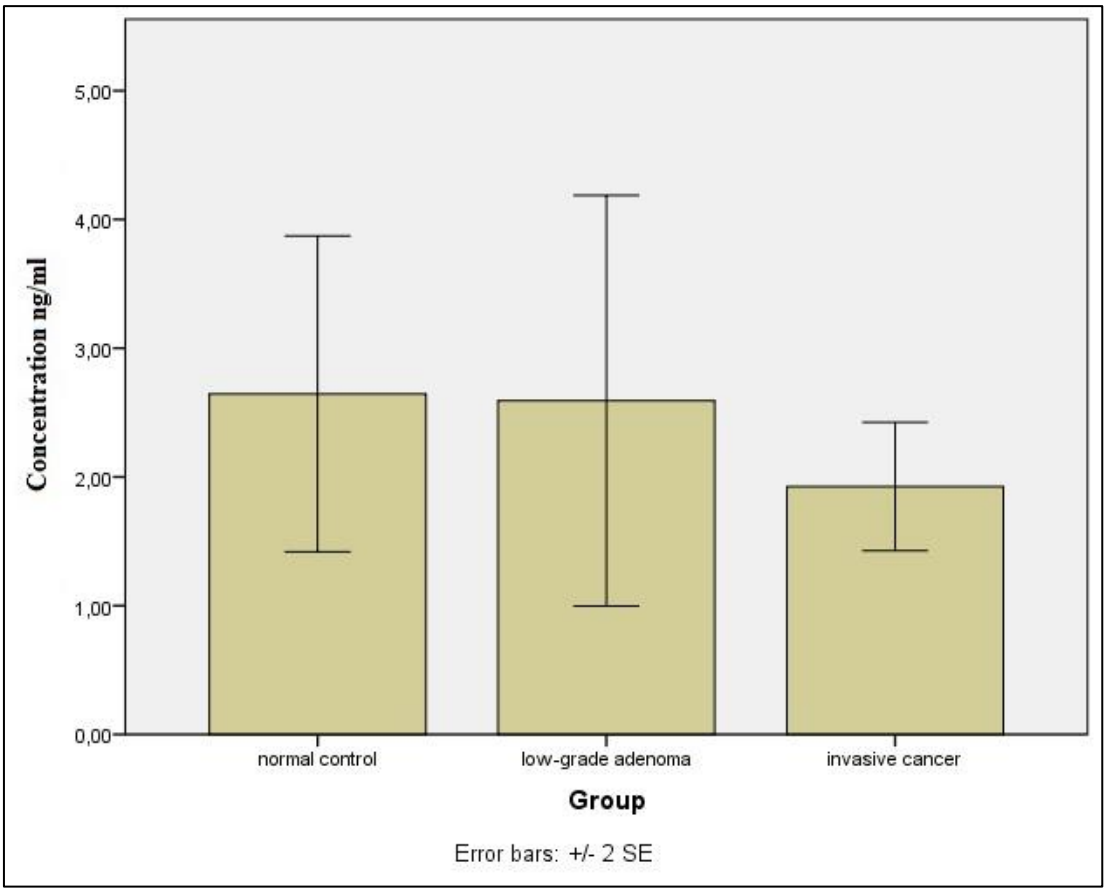


Figure 5.7. Mean serum RPL6 concentration results using the Cloud-Clone ELISA kit. The high-grade adenoma group was excluded from this comparison due to several invalid results. Mean protein concentrations were not statistically different (p=0.396, Kruskal-Wallis test). N: 10 samples per group.

Next, we retested some of the samples used in the Cloud-Clone assay, using an ELISA kit from Abbexa (product number abx570790, Cambridge, UK) to assess the concordance between different kits. In this kit, the protein epitope spans the same area as the Cloud-Clone plate described above. Assay sensitivity, detection range and protocol procedures were identical to the previous kit. As depicted in table 5.9, disregarding the undetectable levels observed in either or both plates, the concordance between results obtained using the two RPL6 kits was superior to the concordance between the different NAP1L1 plates tested earlier. However, discrepancies were still observed and no kit could be deemed an accurate platform. We have not compared results between groups due to the number of invalid readings and the limited number of cases per group.

Table 5.9. Comparison of results obtained with RPL6 ELISA plates from Cloud-Clone and Abbexa.

Sample	CloudClone plate (ng/mL)	Abbexa plate (ng/mL)
N2	3.123	0.914
N4	1.552	1.298
N8	2.286	1.14
N23	0.7	0.384
JJ058	0.378	**Invalid**
JJ093	**Invalid**	**Invalid**
JJ111	0.744	**Invalid**
JJ138	1.364	1.18
JJ156	**Invalid**	**Invalid**
CR62	2.026	2.206
CR65	2.464	2.546
CR70	2.558	2.364
CR73	2.678	2.99
CR76	0.702	**Invalid**

Although the previous experiments have not suggested differences in serum RPL6 concentrations between patient groups, we also tested the expression of this protein in plasma samples, as the optimal matrix for this assessment had not been established. For this analysis, we selected a total of 18 cases (6 normal, 5 high-grade adenomas and 7 cancers) from both the Brazilian and UK cohorts. The criterion for selecting these cases was, again, the availability of sufficient sample volumes for this and additional experiments, if necessary.

Serum and plasma samples from each case were then tested using the Abbexa RPL6 ELISA plate (table 5.10). Similar findings were observed when compared to the results for NAP1L1 plasma-versus-serum assessment (table 5.5, above). Noteworthy, several undetectable levels were obtained for serum samples, whilst all (except one) plasma samples yielded readings within the detection range. Additionally, as occurred with NAP1L1, plasma levels were consistently higher than serum levels – more than 10-fold in some cases, suggesting that plasma rather than serum could again be the optimal biological fluid for RPL6 assessment. Samples from the UK cohort (with the initials JJ) tended to produce higher values than the Brazilian

samples, again suggesting that sample processing or storage conditions may influence protein concentration. Consequently, we focused our attention on the analysis of a larger number of plasma samples as the final strategy in the evaluation of RPL6 as a potential CRC blood biomarker.

Table 5.10. Comparison of RPL6 results between plasma and serum samples using an Abbexa ELISA kit. Plasma samples exhibited higher protein concentrations than serum samples for individual cases.

	Plasma		**Serum**	
Groups	**Samples**	**(ng/mL)**	**Samples**	**(ng/mL)**
Normal Brazil	N2P	1.538	N2S	0.914
	N4P	1.754	N4S	1.298
	N8P	4.068	N8S	1.14
	N23P	2.75	N23S	0.384
Normal Chester	JJ001P	3.264	JJ001S	0.536
	JJ002P	13.786	JJ002S	0.632
High-grade Chester	JJ058P	16.762	JJ058S	**Invalid**
	JJ093P	14.156	JJ093S	**Invalid**
	JJ111P	10.816	JJ111S	**Invalid**
	JJ138P	14.082	JJ138S	1.18
	JJ156P	16.368	JJ156S	**Invalid**
Cancer Brazil	CR62P	5.548	CR62S	2.206
	CR65P	6.798	CR65S	2.546
	CR70P	8.648	CR70S	2.364
	CR73P	**Invalid**	CR73S	2.99
	CR76P	3.12	CR76S	**Invalid**
Cancer Chester	JJ079P	5.14	JJ079S	**Invalid**
	JJ121P	11.026	JJ121S	**Invalid**

Note: the aim of this experiment was solely to compare RPL6 concentrations between plasma and serum samples. No comparison between groups was performed.

As we did for NAP1L1, we analysed RPL6 concentrations in plasma samples from Brazil and from the UK separately in order to further investigate whether different handling or storage conditions had an effect on protein concentrations. Abbexa RPL6 ELISA plates were used. Again, we tested plasma from normal controls and cancer patients for the Brazilian cohort. As illustrated in table 5.11, all samples produced valid results. The mean plasma RPL6 concentration between groups was similar (figure 5.8). In general

however, a wide variation was observed within each group (6- to 8-fold) highlighting the heterogeneity of protein expression within these groups.

Table 5.11. Abbexa RPL6 ELISA kit results using plasma samples from the Brazil cohort.

NORMAL (n=10) Sample Id	Concentration (ng/mL)	CANCER (n=30) Sample Id	Concentration (ng/mL)
N1P	3.35	CR60P	6.40
N2P	2.57	CR62P	4.77
N3P	3.06	CR63P	5.07
N4P	2.26	CR64P	6.97
N5P	2.20	CR65P	5.03
N6P	3.08	CR67P	4.00
N7P	3.30	CR68P	3.64
N8P	3.85	CR69P	6.16
N11P	12.20	CR70P	7.87
N28P	14.97	CR71P	3.75
		CR72P	4.97
		CR73P	1.66
		CR74P	4.10
		CR75P	6.58
		CR76P	2.16
		CR77P	4.32
		CR78P	4.58
		CR79P	6.44
		CR97P	2.36
		CR81P	5.72
		CR82P	8.54
		CR84P	8.62
		CR85P	7.21
		CR86P	13.03
		CR87P	5.27
		CR88P	12.56
		CR89P	2.58
		CR90P	3.97
		CR91P	2.93
		CR92P	5.22

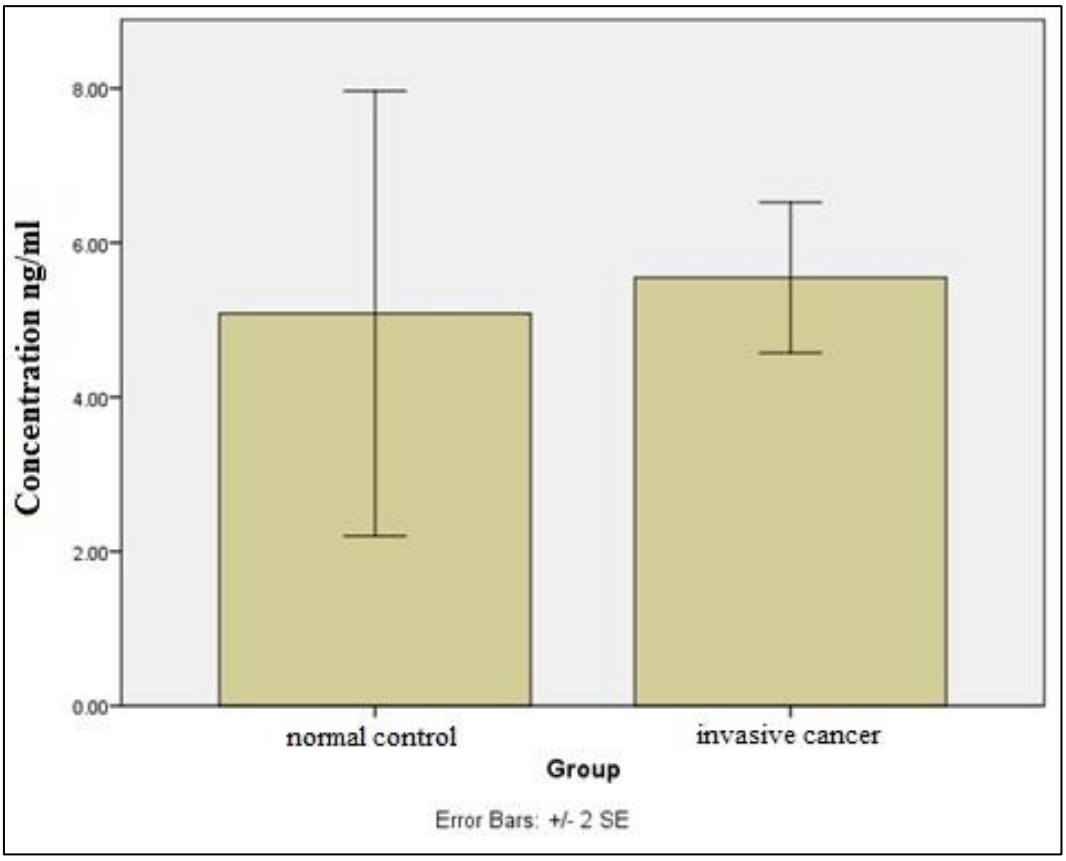


Figure 5.8. Mean RPL6 concentration in plasma samples from the Brazilian cohort (Abbexa kit). No significant difference was observed between groups (p=0.072, Mann-Whitney U test) N: 10 normal controls, 30 invasive cancers.

Next, we performed the same analysis using the UK samples. Plasma from 40 cases encompassing all four groups was tested using the Abbexa plate (table 5.12). Again, plate sensitivity proved to be adequate as valid results were obtained for all samples. However, mean plasma RPL6 concentration was not statistically different between groups (figure 5.9). Noteworthy, mean concentrations for the UK cohort were higher than those observed for the Brazilian cohort. Similarly to NAP1L1, plasma RPL6 concentration appeared to be affected by storage time or some other unclear factor related to sample collection or processing.

Table 5.12. Abbexa RPL6 ELISA kit results using plasma samples from the UK cohort.

Normal (n=10)		Low-grade (n=10)		High-grade (n=7)		Cancer (n=13)	
Sample Id	(ng/mL)	Sample Id	(ng/mL)	Sample Id	(ng/mL)	Sample Id	(ng/mL)
JJ001	4.02	JJ005	4.11	JJ047	3.09	JJ057	14.42
JJ002	20.70	JJ006	12.95	JJ058	20.38	JJ079	6.18
JJ007	8.72	JJ010	10.98	JJ093	17.36	JJ113	10.40
JJ012	14.69	JJ011	5.16	JJ111	11.64	JJ121	11.83
JJ013	30.45	JJ019	13.60	JJ138	14.72	JJ126	9.89
JJ015	27.62	JJ021	50.51	JJ156	23.58	JJ132	31.58
JJ016	3.74	JJ022	21.49	JJ165	3.38	JJ136	25.85
JJ018	7.47	JJ023	20.11			JJ143	23.72
JJ020	69.38	JJ041	2.76			JJ144	21.91
JJ024	23.99	JJ050	18.73			JJ145	20.03
						JJ149	37.19
						JJ157	11.07
						JJ166	4.58

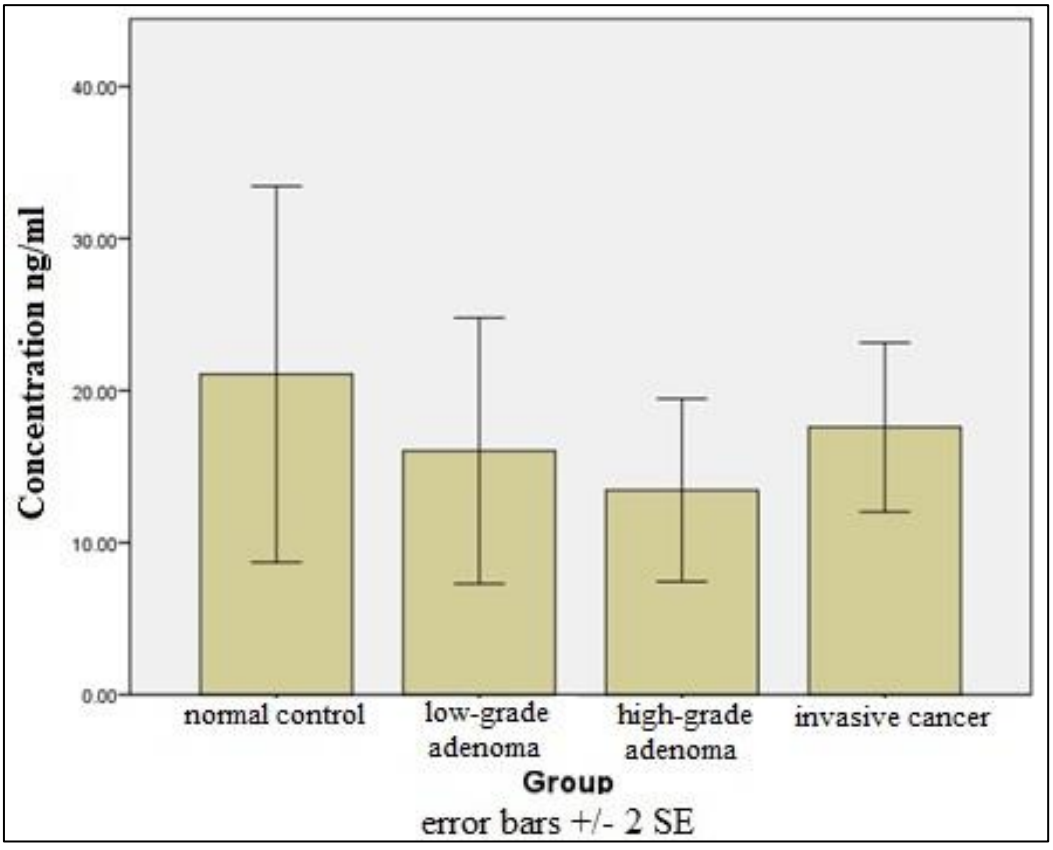


Figure 5.9. Mean RPL6 concentration in plasma samples from the UK cohort (Abbexa kit). Differences observed between groups were not statistically significant (p=0.743, Kruskal-Wallis test). Large intra-group variation was observed. N=10 normal controls, 10 low-grade adenomas, 7 high-grade adenomas and 13 cancers.

In order to obtain a final assessment of the inter-plate consistency of the RPL6 ELISA kit used in most of our experiments (Abbexa kit), we compared samples used in the plasma versus serum plate (table 5.10) which were retested in plasma-only experiments later (tables 5.11 and 5.12). Sixteen plasma samples were used in both experiments. Results produced for each sample by the different plates were reasonably similar (coefficient of variation of 17.7%) and highly correlated (a *Pearson´s* correlation coefficient $r = 0.96$), thus confirming that this RPL6 kit is highly consistent.

These findings suggest that the RPL6 ELISA kits tested were adequate tools for the assessment of the concentration of this protein in biological fluids. Therefore, we can confidently conclude that RPL6 blood levels are not significantly different when comparing normal subjects, individuals with low-grade and high-grade adenomas, and CRC patients. This conclusion is based on the assessment of two different sample cohorts and two biological matrices, tested with ELISA plates from two different providers. Consequently, we decided to terminate the investigation of RPL6 as a potential blood-based biomarker for CRC at this stage. However, some interesting findings were made. Firstly, plasma levels were consistently higher than serum levels. This finding has implications for any future research evaluating RPL6 expression in blood samples. Secondly, in general, the UK cohort exhibited higher detectable concentrations of this protein than the Brazilian cohort. This observation suggests that the length of storage or some other factor during sample collection or processing affects plasma RPL6 concentrations. NAP1L1 exhibited a similar pattern. Similarities between NAP1L1 and RPL6 were also seen during the IHC and qPCR work in this research project. It raises the hypothesis that these two proteins might somehow interact in their biological activities, a possibility that requires further investigation.

5.4.3. **PHB kits – testing and results**

The assessment of PHB as a potential blood biomarker followed the same initial steps described in previous sections. The Cloud-Clone Corp ELISA plate used in the preliminary phase was again tested (product number SEA442Hu, detection range 0.312-20ng/mL). Forty serum samples representing the four clinical groups were initially analysed. Just as we observed with NAP1L1 and RPL6, many samples exhibited undetectable PHB levels, thus preventing any comparison or conclusion (individual data not shown to avoid unnecessary repetition).

To assess whether plasma would also be a better matrix than serum for measuring PHB concentrations, we again tested both fluids using an Abbexa ELISA kit (48-well trial plate, product number abx152835, Cambridge, UK). Twelve specimens (6 serum and 6 plasma samples) were analysed. As depicted in table 5.13, once more, protein concentration in plasma was consistently higher than that observed in serum. One of the serum samples resulted in readings below the lower limit of detection of the kit. Based on these findings, we decided to proceed with this evaluation using plasma samples only. As the detection range of the Abbexa plate seemed to be adequate for this matrix, this plate was chosen for subsequent tests.

Table 5.13. Comparison of PHB results between plasma and serum samples using an Abbexa ELISA kit. Again, plasma samples exhibited higher protein concentrations than serum samples for each individual case.

Plasma		**Serum**	
Samples	**(ng/mL)**	**Samples**	**(ng/mL)**
JJ138P	16.53	JJ138S	1.918
JJ156P	11.392	JJ156S	1.80
JJ057P	11.784	JJ057S	1.287
JJ113P	12.102	JJ113S	1.816
JJ126P	20.401	JJ126S	**Invalid**
JJ143P	15.352	JJ143S	1.96

Similarly to the previous experiments, we tested samples from the Brazilian and UK cohorts separately: normal controls versus cancer samples for the first cohort and all four groups for the second cohort. The analysis of the Brazil cohort is shown in table 5.14 and figure 5.10. Valid results were obtained for all samples, except one normal control subject. Compared to NAP1L1 and RPL6, PHB concentration exhibited less variation between samples resulting in a narrower standard error, and suggesting a more stable protein level. The comparison between normal and cancer groups however showed no difference in PHB concentration.

Table 5.14. Abbexa PHB ELISA results using plasma samples from the Brazilian cohort.

NORMAL (n=10)	Sample Id	Concentration (ng/mL)	CANCER (n=30)	Sample Id	Concentration (ng/mL)
	N1P	8.27		CR60P	3.56
	N2P	5.06		CR62P	8.93
	N3P	7.08		CR63P	4.69
	N4P	5.88		CR64P	9.67
	N5P	5.67		CR65P	6.59
	N6P	11.04		CR67P	6.32
	N7P	2.27		CR68P	5.81
	N8P	6.71		CR69P	10.73
	N11P	**Invalid**		CR70P	8.66
	N28P	8.52		CR71P	8.28
				CR72P	7.04
				CR74P	4.33
				CR75P	10.76
				CR76P	12.32
				CR77P	3.65
				CR78P	12.38
				CR79P	11.89
				CR97P	6.58
				CR81P	8.56
				CR82P	9.74
				CR84P	11.18
				CR85P	7.89
				CR86P	9.52
				CR87P	8.66
				CR88P	6.70
				CR89P	8.03
				CR90P	11.60
				CR91P	9.83
				CR92P	8.04

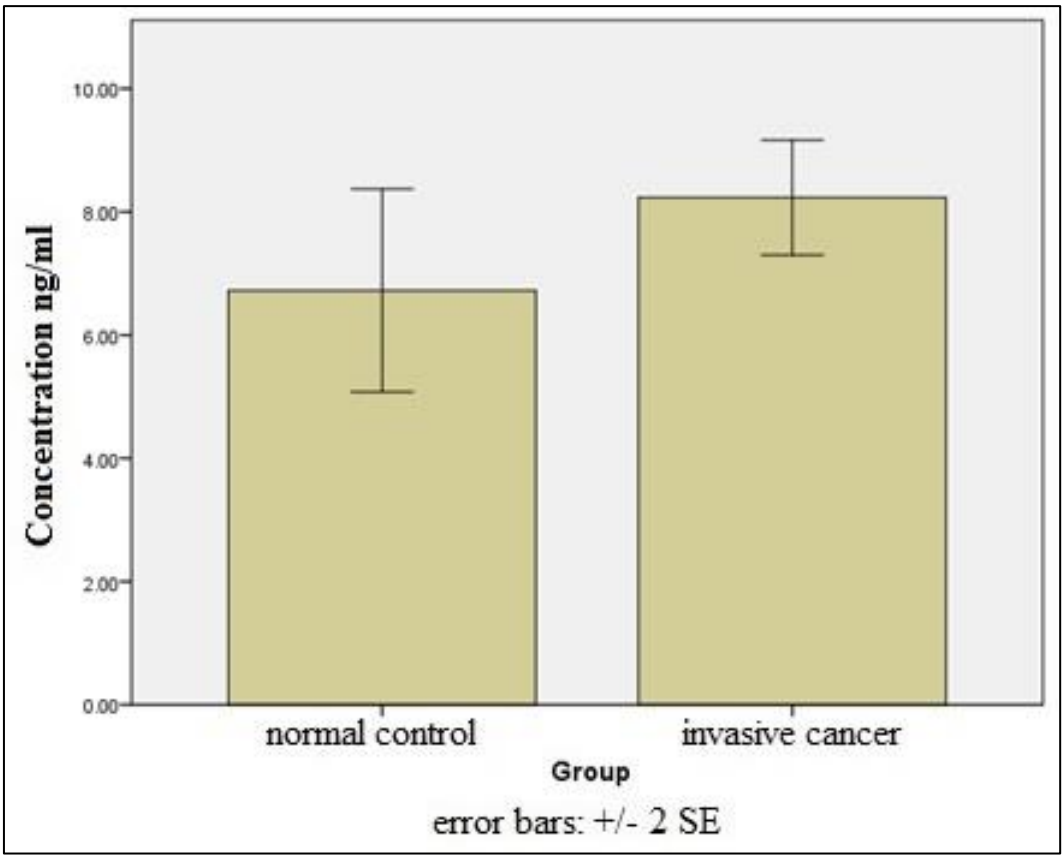


Figure 5.10. Mean PHB concentration in plasma samples from the Brazilian cohort (Abbexa kit). No significant difference was observed between groups (p=0.619, Mann-Whitney U test) N: 10 normal controls, 30 invasive cancers.

For the UK cohort, all samples tested produced valid results (see table 5.15). Again, PHB concentrations were relatively homogeneous, except for a few outliers. A comparison of the mean PHB concentration between the four clinical groups however did not reveal any significant difference (figure 5.11). Five samples were tested twice in separate experiments using Abbexa plates. The calculated inter-assay coefficient of variation for these samples was 11.1%.

Table 5.15. Abbexa PHB ELISA kit results using plasma samples from the UK cohort.

Normal (n=10)		Low-grade (n=10)		High-grade (n=7)		Cancer (n=13)	
Sample Id	(ng/mL)	Sample Id	(ng/mL)	Sample Id	(ng/mL)	Sample Id	(ng/mL)
JJ001	15.71	JJ005	1.12	JJ047	11.26	JJ057	12.70
JJ002	13.39	JJ006	15.30	JJ058	17.64	JJ079	10.03
JJ007	12.47	JJ010	12.65	JJ093	116.39	JJ113	2.27
JJ012	15.44	JJ011	9.29	JJ111	17.22	JJ121	15.99
JJ013	19.81	JJ019	12.05	JJ138	21.22	JJ126	14.48
JJ015	16.16	JJ021	14.11	JJ156	12.43	JJ132	20.04
JJ016	15.16	JJ022	13.84	JJ165	10.48	JJ136	17.74
JJ018	24.98	JJ023	15.89			JJ143	14.82
JJ020	15.99	JJ041	16.88			JJ144	15.28
JJ024	11.073	JJ050	15.51			JJ145	13.74
						JJ149	18.57
						JJ157	16.73
						JJ166	11.83

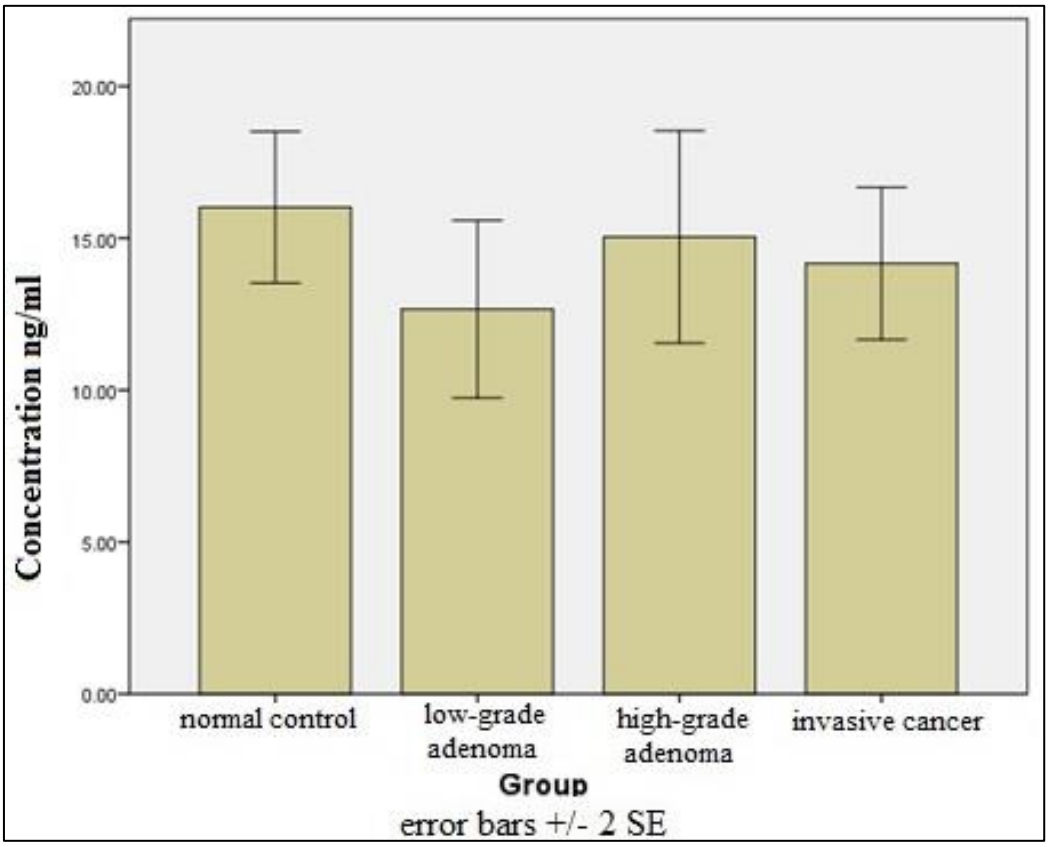


Figure 5.11. Mean PHB concentration in plasma samples from the UK cohort (Abbexa kit). Differences observed between groups were not statistically significant (p=0.593, Kruskal-Wallis test). N=10 normal controls, 10 low-grade adenomas, 7 high-grade adenomas and 13 cancers.

Collectively, these results suggest that PHB blood concentration is altered neither in CRC patients nor in individuals with pre-neoplastic colonic lesions when compared with normal controls. The adequate performance demonstrated by the ELISA kits tested discouraged any further assessment of this protein as a potential CRC blood biomarker, as it is unlikely that a different platform would yield different results. However, it must be emphasised that, as for NAP1L1 and RPL6, PHB concentration in plasma was consistently higher than in serum. This finding suggests that the former is a better matrix than the latter for future studies assessing PHB blood concentrations.

5.5. NAP1L1 electrochemiluminescence assay development

As described in section 5.4.1, the detailed assessment of NAP1L1 as a potential CRC blood biomarker was impaired by the lack of sensitivity and consistency of the ELISA kits tested. Only a non-significant trend of increased protein concentration in the plasma of patients with high-grade adenomas and CRC was observed, but the general performance of the kits was poor. In order to definitively confirm or refute whether NAP1L1 has potential as a CRC blood biomarker, we decided to develop an in-house immunoassay which would hopefully have better performance and reliability than the commercial ELISA kits. As chemiluminescence methods are reportedly more sensitive than enzymatic techniques (Price and Newman, 1997, Blackburn *et al*., 1991), Meso Scale Discovery® (MSD) – electrochemiluminescence (ECL) was the chosen platform. MSD performance has previously been specifically compared with other immunoassays. As demonstrated in table 5.16, Leary *et al*. showed that MSD is superior to ELISA and comparable to bead-based methods for the assessment of generic human IgG. A wider dynamic range and lower limit of detection were observed compared with the enzymatic method. Other advantages of the MSD platform are the need of lower sample volumes (usually 25-30µL of final volume per well), shorter procedure time and the

potential for multiplexing, thus allowing the evaluation of up to 10 markers in one microtiter plate (Leary *et al*., 2013, Marchese *et al*., 2009).

Table 5.16. Comparison of performances between selected immunoassays targeting a generic human IgG. Adapted from (Leary *et al*., 2013)

Platform	**Format**	**Readout**	**Dynamic range (ng/mL)**	**Sensitivity (ng/mL)**
ELISA	Plate	Colorimetric	88.0-666	88.0
MSD	Plate	ECL	15.6-4000	15.6
Gyrolab	Bead	Fluorescence	10.5-6400	10.5
AlphaLISA	Bead	Luminescence	181-1097	181.0

MSD technology has proven to be superior to ELISA for the assessment of various proteins such as neurofilament light chains (Kuhle *et al*., 2016) and ricin B chain (Guglielmo-Viret and Thullier, 2007). An in-house MSD assay produced by Chaturvedi *et al*. was superior to ELISAs and even to a commercial MSD kit for the assessment of total interleukin-6, one of the least expressed proteins in blood samples (Chaturvedi *et al*., 2015). Marchese *et al*. developed a multiplex MSD assay for the concomitant assessment of eight anti-pneumococcal antibodies related to the vaccine Pneumovax 23, and this outperformed the use of standard individual ELISA tests for each antibody (Marchese *et al*., 2009). Therefore, the MSD-ECL platform was a suitable method to try to develop a sensitive and consistent assay for NAP1L1 measurement.

5.5.1. **Principles**

Sandwich-format MSD-ECL follows the same principles of other immunoassays such as ELISAs. A microtiter polystyrene plate is coated with a capture reagent which binds to the plate via hydrophobic interactions. The capture substance may be a protein, a carbohydrate, viral particles or cell components, among others (figure 5.12, upper panel). In a classical sandwich immunoassay, a capture antibody is used to target specific analytes (figure 5.12, lower panel). After the initial coating, each well undergoes a blocking step – incubation with a protein solution aiming to

occupy any uncoated space in the well, thus reducing non-specific binding during the following phases of the procedure. Different proteins may be used for blocking and the best option should be determined for each immunoassay. After blocking, samples containing the analyte are added to the wells, along with a dilution series of the standard protein. This standard curve is later used to interpolate each sample reading and determine individual sample concentrations. Next, a detection antibody is used to identify analytes bound to the capture reagent. This antibody may be tagged with a reporter substance such as alkaline phosphatase or horseradish peroxidase (in ELISAs), or a fluorescent/luminescent substance. Alternatively, a third anti-species antibody targeting the detection antibody is used as the reporter. Finally, a substrate (ELISAs) or an electrical current (ECL) is applied to the plate, allowing the reporter to react, thus producing colour, fluorescence or light. A microplate reader is then used to record the signal, which is proportional to the concentration of the analyte in each well.

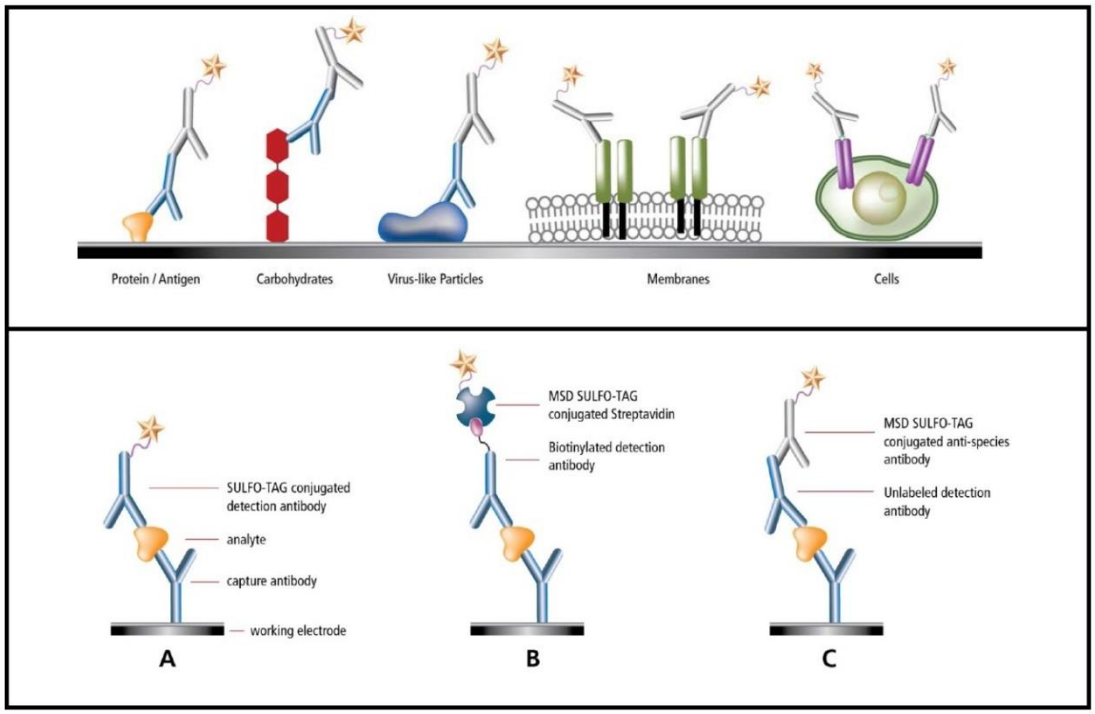


Figure 5.12. Capture reagents and assay formats used in MSD experiments. In the upper panel, different components are used as the capture reagent aiming to attract specific antibodies. In the bottom panel, capture and detection antibodies are used to assess a particular analyte. The reaction is reported using MSD SULFO-TAG™, which can be directly conjugated with the detection antibody (A), attached to it via biotin-streptavidin interaction (B) or linked to a tertiary anti-species antibody (C).

MSD-ECL plates have several adaptations, compared to ELISAs, to allow electricity to be used to stimulate the final reaction. Each well bottom is a working electrode where all the adsorption of capture reagents, analytes and detection reagents occur (figure 5.13). The reporter reagent is an antibody tagged with an electrochemiluminescent label called SULFO-TAG™, a proprietary compound that emits light upon the electrical stimulation provided by the MSD plate reader. The essential chemical component of SULFO-TAG™ is the molecule tris(2,2′-bipyridyl)ruthenium(II) or $Ru(bpy)_3^{2+}$. When the necessary voltage is applied, this compound undergoes a series of oxidative-reductive reactions and interacts with a co-reactant called tri-n-propylamine, or $TPA^{+\cdot}$, present in the Read Buffer. These reactions result in the emission of red light at 620nm and the regeneration of $Ru(bpy)_3^{2+}$ to its ground state thus enhancing the sensitivity of the method (Wei and Wang, 2011). Additionally, the proximity-based nature of the reaction (only luminescent labels in close contact with the electrode will react) reduces non-specific signals in the solution or in the wall of the wells improving the accuracy of the test.

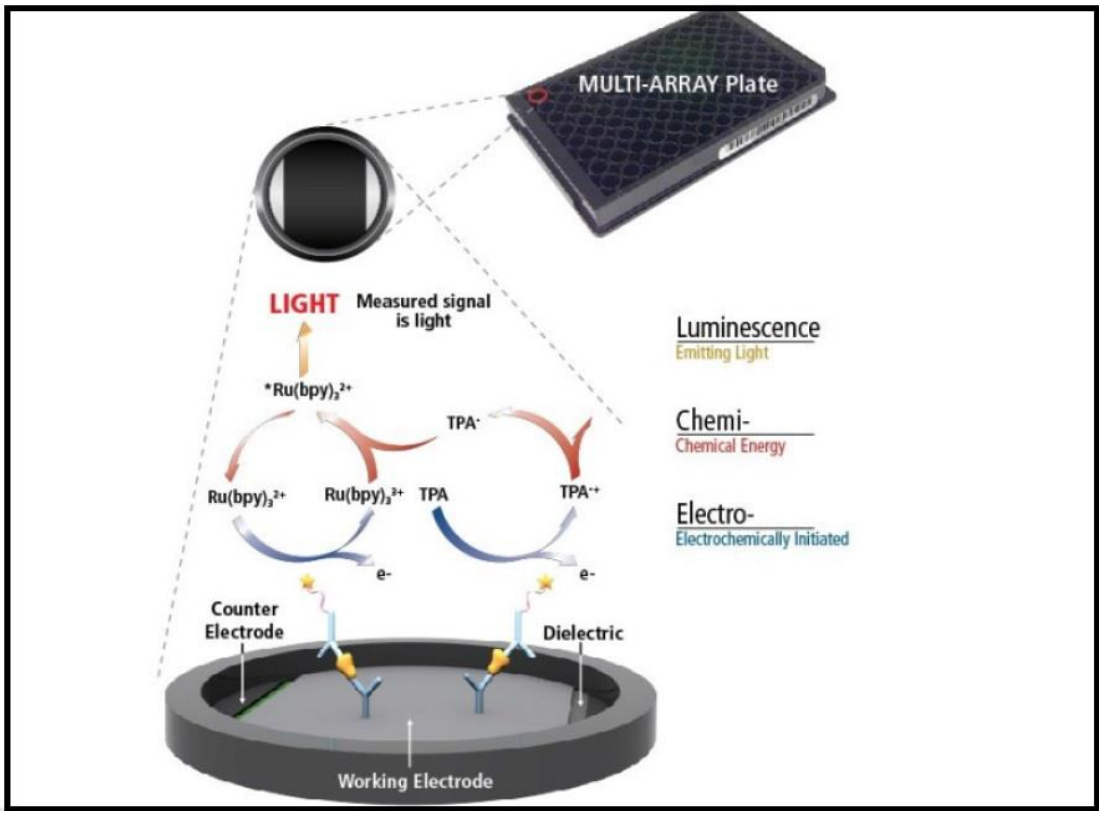


Figure 5.13. Illustrative representation of the MSD-ECL technology.

General procedures suggested by the manufacturer and used in this assay development are:

a. Coat plates with 30µL of capture antibody solution in PBS per well overnight at 4°C;
b. Wash plate 3 times using 200µL of PBS-Tween 0.05%;
c. Add blocking reagent and incubate for 1 hour in a plate shaker at >300rpm, at room temperature;
d. Wash plate 3 times, as above;
e. Add 25µL of samples or standards diluted in blocking solution to the assigned wells and incubate for 1 hour in a plate shaker at >300rpm, at room temperature;
f. Wash plate 3 times, as above;
g. Add 25µL of the detection antibody diluted in blocking solution and incubate for 1 hour in a plate shaker at >300rpm, at room temperature;
h. Wash plate 3 times, as above;
i. Add 25µL of SULFO-TAG™ labelled anti-species antibody at 1.0µg/mL diluted in blocking solution and incubate for 1 hour in a plate shaker at >300rpm;
j. Wash 3 times, as above;
k. Add 150µL of 2x Read Buffer and read plate immediately.

Two types of microtiter plates are produced by the manufacturer: a High Bind (HB) plate in which hydrophilic adsorption of capture reagents occurs and a Standard Bind (SB) plate in which hydrophobic interactions take place. It is recommended that both plate types are tested when developing a new assay. The plate reader used in this research was the proprietary SECTOR Imager 6000 and results were analysed with the MSD Discovery Workbench software, version 4.0. The method of analysis used was the four-parameter logistic (4PL) regression model, with a $1/Y^2$ weighting.

All MSD equipment, plates and specific reagents were manufactured in Rockville, MD, USA.

5.5.2. **Assay development overview**

The development of a novel immunoassay requires the careful testing of all the components involved in the procedure. From antibody selection to the concentration and composition of individual reagents, each detail can affect the result and the validity of an immunoassay (Del Campo *et al*., 2015). The initial step in the assay development process is the design, production and selection of adequate antibodies. When available, commercial antibodies may be tested in order to find appropriate options. The aim is to find a combination of antibodies (capture and detection) that specifically identifies the target protein with the lowest background. When an adequate pair has been identified, the optimal concentration able to produce a satisfactory signal whilst maintaining a low background is pursued. These procedures establish the initial conditions to proceed with the assay development.

The next step is the assay set-up aiming to test both the calibrator and clinical samples. The calibrator (also called reference protein or standard) is a purified protein or peptide that is used to produce a standard curve. Calibrator highest concentration, dilution factor and diluent buffers must all be defined. The aim of this step is to produce an optimal standard curve with a wide dynamic range and good dilution linearity. After obtaining an adequate standard curve, the next stage in assay development is the testing of clinical samples. In this step, the sample dilution factor to be used is defined. Additionally, aberrant results due to the interference provoked by the several constituents of the biological fluid are unveiled, a phenomenon called matrix interference. During all the previous steps, the composition of the blocking buffer, dilution buffer and wash buffer must be tested as these reagents may influence the signal-to-background ratio (SBR), an essential variable for a good immunoassay.

Lastly, after all the optimal conditions have been identified, the novel assay must undergo validation or quality-control experiments. These procedures aim to define assay imprecision (assessed by inter- and intra-assay CV), recovery, dilution linearity, among others (Del Campo *et al*., 2015, Andreasson *et al*., 2015). After all these steps have been successfully

performed, clinical samples can be tested in order to produce final results. An overview of the several experiments that were executed to produce an in-house MSD ECL assay for NAP1L1 is now provided.

5.5.3. **Selection of antibodies, calibrators and plate types**

Anti-NAP1L1 antibodies are currently produced by several manufacturers. Therefore, our first task was the selection of some options for initial antibody screening. In this research, we opted for the use of a tertiary SULFO-TAG™-labelled antibody, as depicted in figure 5.12 above (bottom panel, item C). An essential requirement for the selection of a capture-detection pair is the use of antibodies produced in different species (a mouse-derived capture antibody and a rabbit-derived detection antibody, for example) so that the use of an anti-species tertiary antibody is possible. Another strategy is to test both monoclonal and polyclonal antibodies in order to determine which yield the best results. Thus, we initially selected five anti-human NAP1L1 antibodies. Two of them were monoclonal and the remaining three were polyclonal. Four of them were rabbit-derived and one was mouse-derived. These initial candidates are described in more detail below:

a. Abcam rabbit monoclonal antibody (ab178687, Cambridge, UK) – epitope: aa 300 to the C-terminus. The exact sequence is proprietary;
b. Abcam rabbit polyclonal antibody (ab33076, Cambridge, UK) – epitope: aa1 – 100;
c. Genetex rabbit polyclonal antibody (gtx112613, Irvine, CA, USA) – epitope: centre region of human NAP1L1. The exact sequence is proprietary;
d. Genetex rabbit polyclonal antibody (gtx124370, Irvine, CA, USA) – epitope: not informed;
e. Abnova mouse monoclonal antibody (H00004673-M01, Taipei, Taiwan) – epitope: aa 1-110.

The antibody screening process started with testing the reactivity of these candidates in the ECL platform. For this purpose, both high bind (HB) and standard bind (SB) plates were used. Using the general procedures described in section 5.5.1, the plates were coated overnight with each of the four rabbit anti-NAP1L1 antibodies at 10µg/mL. The next day, plates were incubated with the MSD proprietary Blocker A (a bovine serum albumin blocking solution). We then generated a 4-fold serial dilution of the standard protein which was provided as part of the Cloud-Clone ELISA kit that we tested in section 5.4.1 (initial concentration: 1,000ng/mL). Next, the Abnova mouse monoclonal antibody was added at 10µg/mL as the detection reagent. A SULFO-TAG™-labelled anti-mouse antibody was then used. After the final wash step, 2x Read Buffer was added to the wells and the plate was read using the MSD SECTOR Imager 6000 plate reader. As seen in figure 5.14, different concentrations of the Cloud-Clone standard protein did not show any difference in signal reading, proving that no antigen-antibody reaction had occurred. As we did not know whether this standard was an incomplete protein or a peptide sequence not covered by the antibodies, we decided to test a commercially available full-length protein instead.

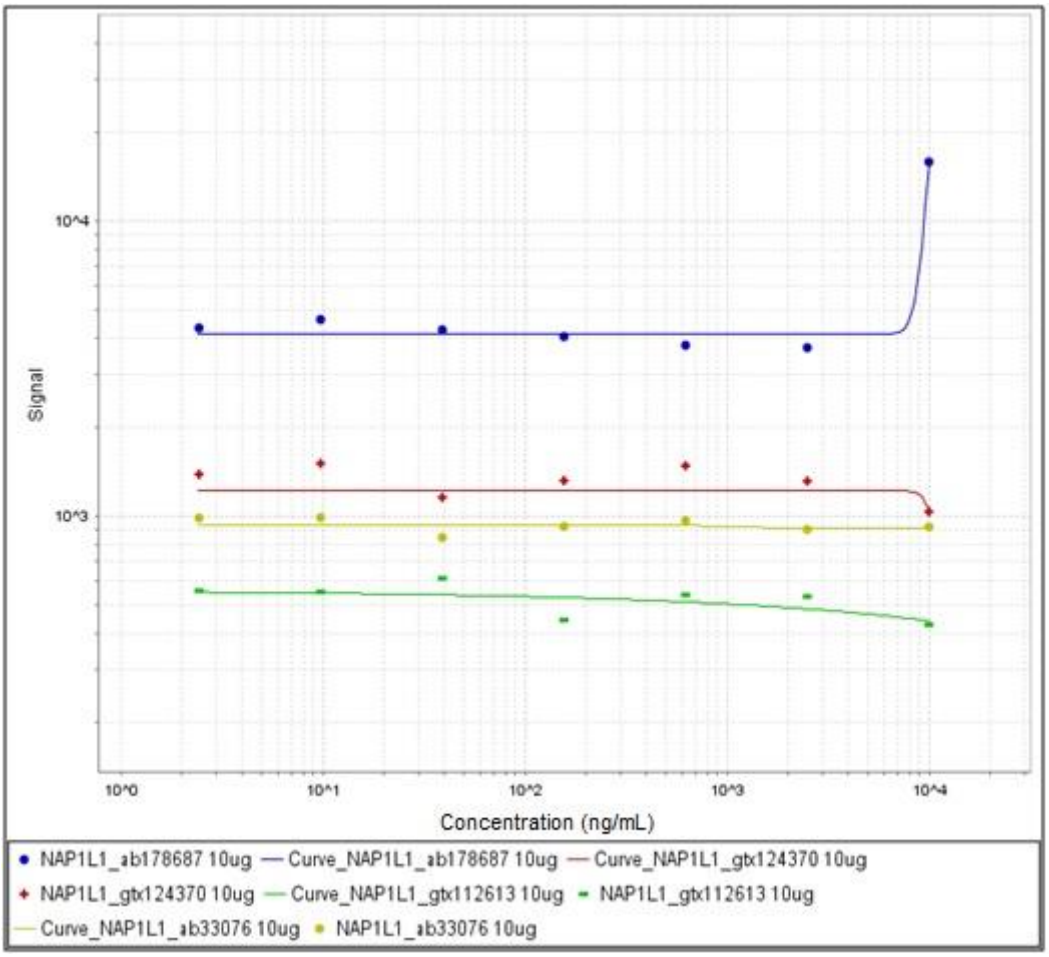


Figure 5.14. Initial screening of antibody pairs using the Cloud-Clone NAP1L1 standard protein as a calibrator. Four different rabbit antibodies were used for capture and the Abnova mouse monoclonal antibody was used for detection. No difference in signal was observed for solutions with different protein concentrations.

For the next experiment, we purchased human NAP1L1 recombinant protein from Abcam (product number ab117213, Cambridge, UK). The procedures used in the last test were replicated. This time, as observed in figure 5.15, a typical standard curve was produced for each of the antibody combinations, thus confirming that this standard protein was properly identified by the antibodies and could be used as the calibrator for the next development steps.

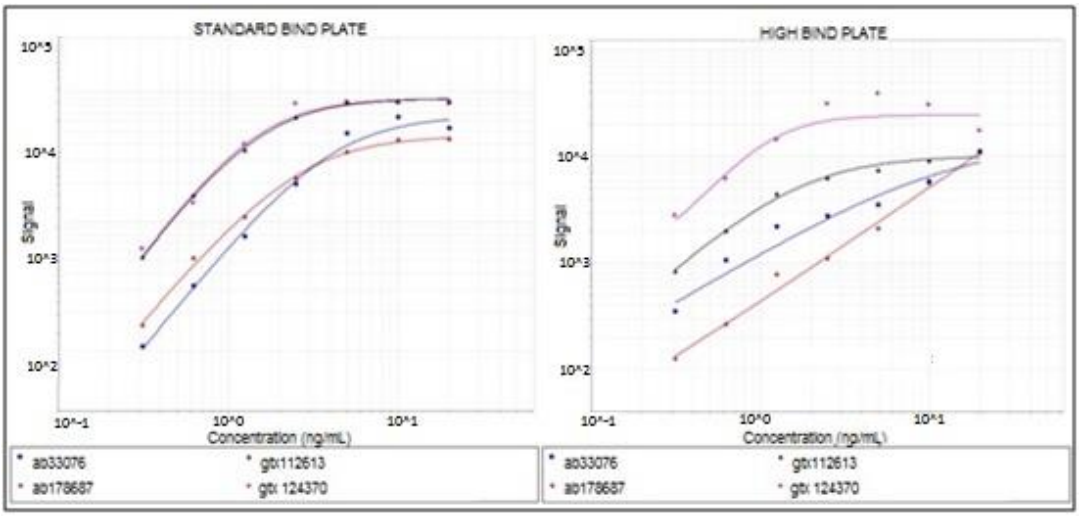


Figure 5.15. The conditions previously used were replicated using the Abcam human recombinant NAP1L1 protein as calibrator. All four antibody pairs exhibited ascending signal readings paralleling increasing protein concentrations. However, an early plateau was observed in the top part of most curves.

A plateau in the top portion of the majority of the standard curves denoted early saturation of the assay. This may be caused by excessive amounts of reagents such as antibodies or calibrator. Therefore, in the next step we tested lower antibody concentrations whilst the maximum calibrator concentration was also reduced to 500ng/mL (in a 2-fold dilution series). In addition, we tried using the Abnova monoclonal mouse antibody as the capture reagent and the rabbit mono and polyclonal antibodies for detection. Capture antibody concentrations were reduced to 2.5 and 5.0µg/mL, and the detection antibody concentration was 1.0µg/mL. Figure 5.16 shows an example of the standard curves that were obtained for one of the conditions tested. Similar curves were observed for each of the other conditions. Therefore, a meticulous assessment of individual readings was necessary to select the best pairing. The initial variable used to screen for adequate antibody conditions in immunoassay development is the signal-to-background ratio (SBR). A high SBR coupled with a steady increase in signal from “zero” to top standard concentration usually translates into a wide dynamic range. In figure 5.17, readings (in arbitrary units) for all conditions tested are shown. The use of monoclonal antibodies (Abnova mouse and Abcam rabbit, respectively) as capture and detection reagents yielded the highest background (in green), resulting in low SBRs. Therefore, this

combination pair was promptly excluded from further experiments. From the remaining conditions, one antibody pair produced very high SBRs in each plate types: Abnova mouse monoclonal – Gtx rabbit polyclonal in the HB plate, and Abnova mouse monoclonal – Abcam rabbit polyclonal in the SB plate (highlighted with red boxes in the figure). Noteworthy, the conditions using the lower capture antibody concentration (2.5µg/mL) demonstrated a better performance. Thus, these two antibody pairs were further tested in the subsequent experiments.

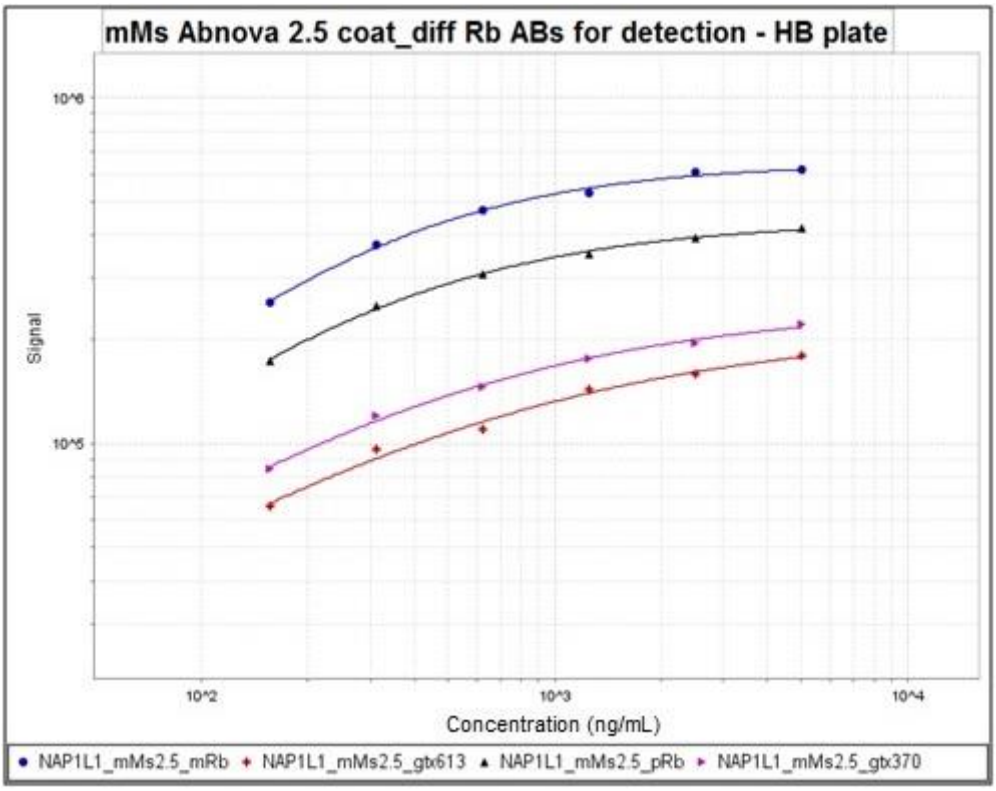


Figure 5.16. An example of the standard curves produced when testing the Abnova mouse monoclonal antibody for capture (in this case, at 2.5µg/mL) and the four rabbit antibodies for detection (at 1.0µg/mL).

Plate type and code	Sample Group	0	156.25	312.5	625	1250	2500	5000	SBR
HB_Plate_*22B13A9339A*	mMs2.5_mRb	28087	255452	375138	472218	530100	608245	618779	22
	mMs5.0_mRb	31315	274901	401698	544613	594855	648122	656826	21
	mMs2.5_pRb	312	173355	250184	308488	352912	393186	419521	1345
	mMs5.0_pRb	389	204394	312444	400253	442358	481161	496398	1276
	mMs2.5_gtx370	165	84346	120229	145844	176079	195016	221259	1341
	mMs5.0_gtx370	270	90379	141015	178997	204418	220416	246787	914
	mMs2.5_gtx613	102	65853	96136	109771	143213	158709	179296	1758
	mMs5.0_gtx613	144	73027	117622	148175	167226	181338	186589	1296
SB_Plate_*21C2RA2159S*	mMs2.5_mRb	3327	112592	153887	216769	276658	374108	547389	165
	mMs5.0_mRb	3948	171950	248422	293697	345126	476321	590356	150
	mMs2.5_pRb	286	145942	207428	250649	296231	374856	495528	1733
	mMs5.0_pRb	853	209102	308012	379692	442855	493191	562142	659
	mMs2.5_gtx370	275	36527	45053	59159	71144	81910	123185	448
	mMs5.0_gtx370	1057	30959	70812	85224	93403	101811	134365	127
	mMs2.5_gtx613	287	29671	42407	44873	55698	69536	90330	315
	mMs5.0_gtx613	564	40439	43215	58668	61969	82614	94902	168

Figure 5.17. Standard curve readings for each of the conditions tested in both HB and SB plates. Signal-to-background ratios (SBR) are used to select the best performing antibody pairs and concentrations. The best conditions in each plate type are circled with red boxes.

Despite these promising results, the signal increase was not entirely proportional to the increase in standard concentration, resulting in flat standard curves (above). Again, we hypothesised that an excessive concentration of reagents was interfering with the performance of the test. Therefore, for the next stage of assay development, lower capture antibody concentrations were added (2.5 and 1.0µg/mL) and the top standard concentration was reduced to 100ng/mL. Additionally, the standard curve was produced using a 3-fold dilution series in order to better explore the lower part of the dynamic range. These assay formats also produced good SBRs (data not shown) and, more interestingly, demonstrated an increase in signal that was proportional to the different concentrations in the standard curve, as can be noticed by the slope of the standard curves in figure 5.18. However, once more a plateau was observed in the top portion of the curve without any flattening in the bottom part. Given the very low protein levels observed in the ELISA experiments described earlier, a further exploration of the lower part of the dynamic range was recommended.

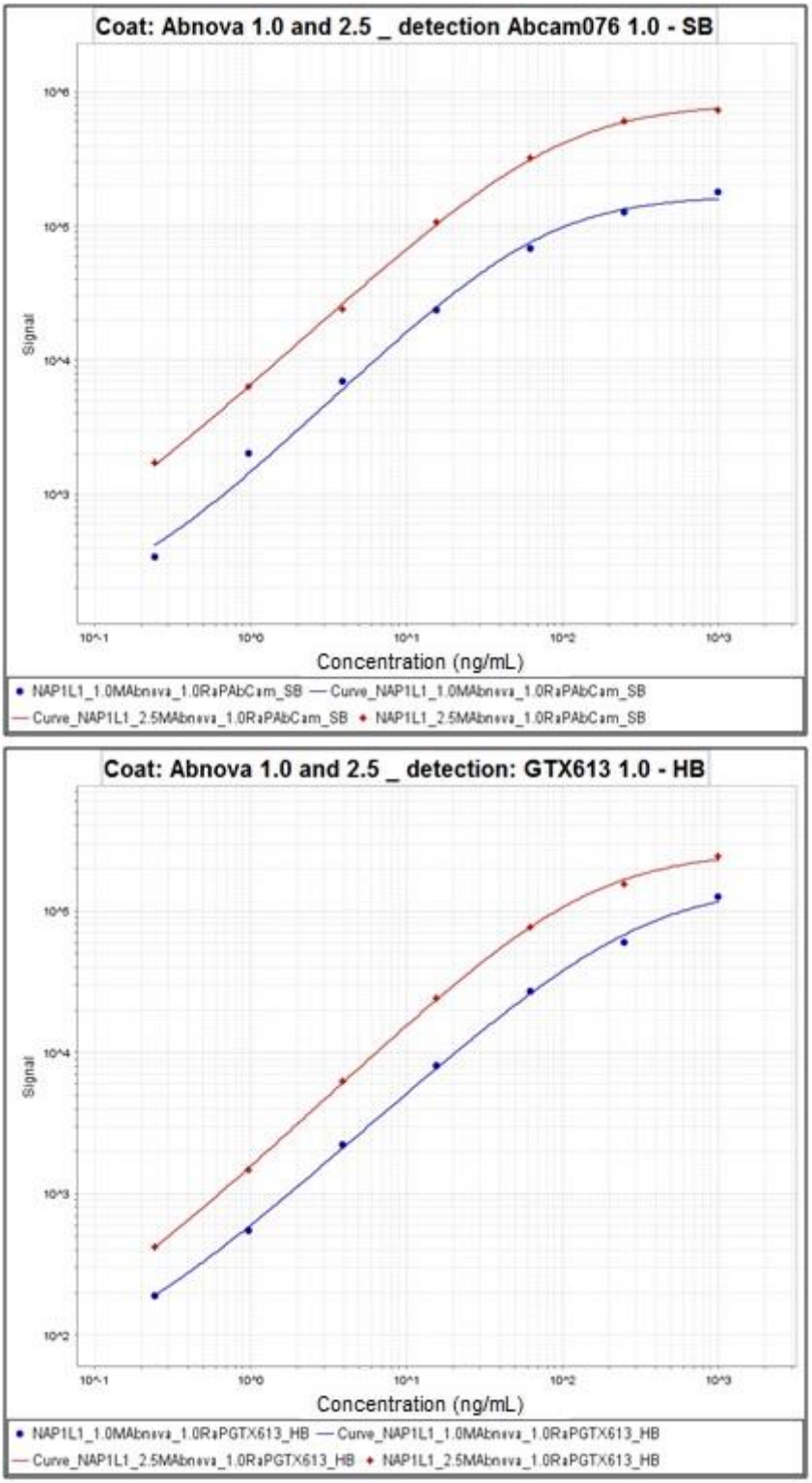


Figure 5.18. The best antibody pairs and conditions previously identified are tested in the respective plate types (SB in the upper graph, HB in the lower graph). The Abnova antibody was tested at both 1.0 (blue line) and 2.5µg/mL (red line) as the capture reagent. Detection antibodies were used at 1.0µg/mL.

From the two conditions selected, we decided to proceed with assay development using the Abnova mouse monoclonal – Abcam rabbit polyclonal antibody pair. This option was chosen because this pair exhibited the best performance in the SB plate. As the name suggests, SB plates have hydrophobic surfaces with binding capacity similar to regular microtiter plates used in other immunoassays, such as ELISAs. This would theoretically allow the transfer of our in-house assay conditions to other platforms, when optimised. Consequently, we next tested this antibody pair in an assay using a further reduced top standard concentration (62ng/mL), a capture antibody concentration of 2.5µg/mL and a detection antibody concentration of 1.0µg/mL. In addition, the inverted format (Abcam antibody as capture and Abnova antibody as detection) was also tested alongside. An additional experiment using the same conditions as above, but using a 4-fold dilution series for the standard curve was performed. This last condition is depicted in figure 5.19. As can be observed, the standard curve exhibited an appropriate format, although a clear flattening was seen in its lower part, suggesting that the 4-fold dilution was probably unnecessary.

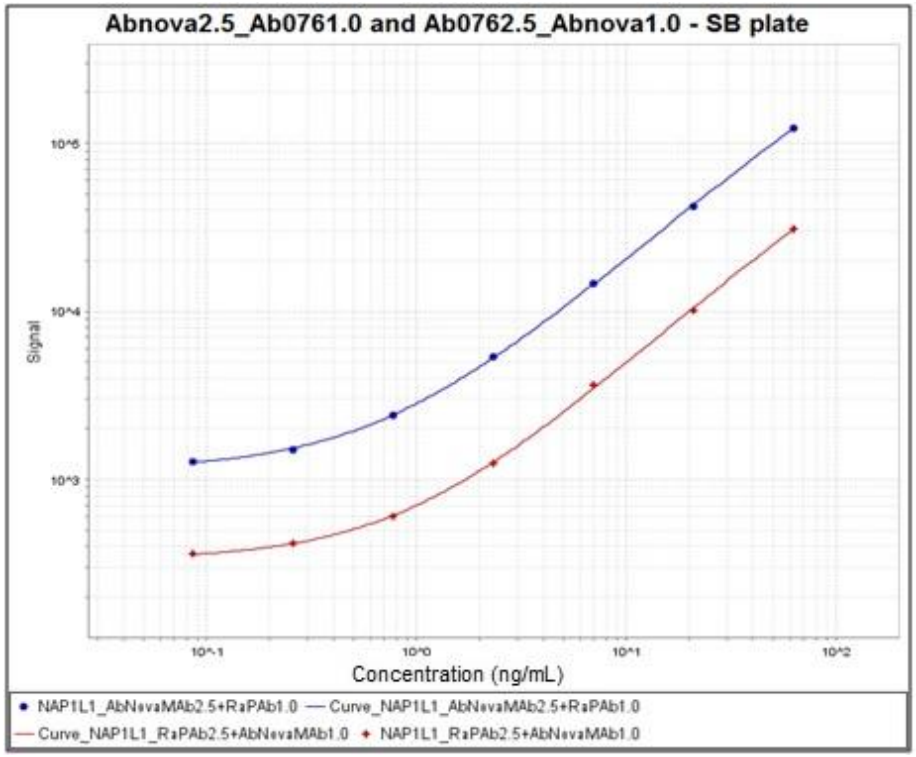


Figure 5.19. Experiment testing the optimal antibody pair and the selected plate type (SB). Abnova mouse monoclonal and Abcam rabbit polyclonal antibodies were used for capture-detection (blue line) and in the opposite orientation (red line).

The experiments previously described demonstrated that the Abnova mouse monoclonal antibody (H00004673-M01) at 2.5µg/mL for capture along with the Abcam rabbit polyclonal antibody (ab33076) at 1.0µg/mL for detection produced the best results when the assay was performed in SB plates. Additionally, the Abcam human recombinant NAP1L1 protein (ab117213) proved to be an adequate calibrator for the standard curve. The highest concentration of 62ng/mL and the 3-fold dilution series resulted in the best standard curve. The use of these conditions produced the steady and consistent curve depicted in figure 5.20. With these basic assay conditions established, we proceeded to assess clinical samples in order to define how the matrices interfere with the results and what sample dilution minimises these effects.

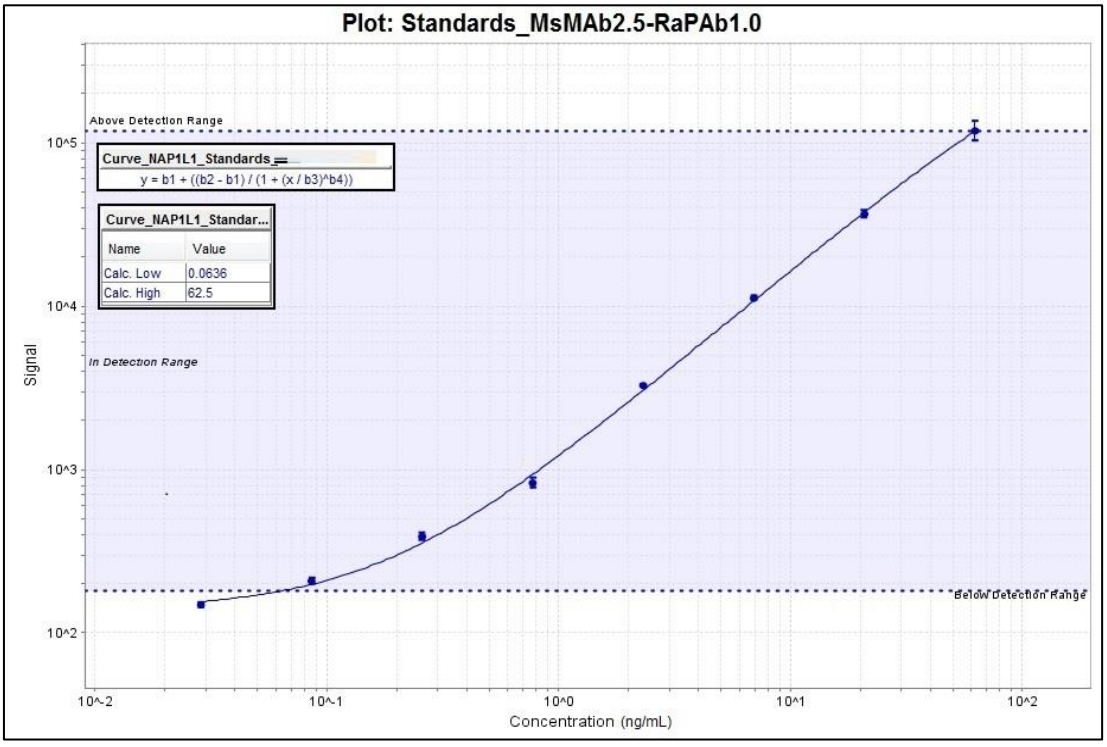


Figure 5.20. Standard curve produced by the selected assay conditions. The dynamic range (0.06 – 62.5ng/mL) spans 3 logs, outperforming the ELISA kits previously tested. This result established the basis for the next phases of assay development.

5.5.4. **Sample dilution and matrix of choice**

There are two main reasons for diluting samples before performing immunoassays: firstly, to reassure that the concentration lies within the dynamic range of the test (when concentrations are expected to be very high) and, secondly, to minimise any interfering effect from the matrix. The human blood proteome is made up of thousands of proteins, either in their native form or as several variations (including precursors, “mature” forms, degradation products, splice variants, glycosylated forms, etc) (Anderson and Anderson, 2002). Several logs of dynamic ranges are encompassed, from albumin (range: 35-50 mg/mL) to interleukin-6 (range: 0-5 pg/mL). Therefore, the study of any low-expressed protein is challenged by the multitude of other bioactive substances that can affect the accuracy of any test.

In order to disclose matrix interference, the first test performed in this research was the assessment of dilution linearity and initial spike recovery. For this purpose, two serum samples (one from a cancer patient and one from a healthy volunteer) were tested in undiluted form, and diluted 1:2, 1:3 and 1:4. If no matrix interference takes place, a proportional reduction in sample signal is expected. On the other hand, in the case of disproportional readings, the dilution range exhibiting concentrations closest to the expected values should be selected. Additionally, each sample (undiluted and diluted) was spiked with known amounts of the calibrator. In low-spiked samples, an increase of 3.1ng/mL was expected, whereas in high-spiked samples, an increase of 6.2ng/mL should occur. In the absence of any interference, 100% recovery of the spiked protein was expected. Ideally, dilution linearity and spike recovery should be maintained between 80-120% of the expected value (Andreasson *et al*., 2015, Leary *et al*., 2013). In this initial test, conditions in which these parameters were between 75-125% were selected. Results are demonstrated in table 5.17. The best performing conditions are highlighted in bold. As can be seen, different results were obtained for serum samples from cancer and normal individuals. For the cancer serum sample, both 1:2 and 1:3 dilutions showed poor linearity, whilst 1:4 exhibited an adequate correlation. Similarly, spike recovery for this sample was optimal for

the sample diluted 1:4, and inadequate for both the neat sample and the other dilutions. These findings suggest that 1:4 is probably the best sample dilution factor for this marker/matrix. The results of dilution linearity for the healthy volunteer serum were below the lower limit of detection, thus preventing us from drawing any conclusions. Assuming a concentration of zero (which underestimates the expected concentration), spike recovery was low for all dilutions. It is unclear whether this is the result of the non-spiked samples being below the detection range or due to an intrinsic factor related to this individual sample.

Table 5.17. Initial tests of dilution linearity and spike recovery using serum samples. Results within 75-125% of the expected value are highlighted in bold.

	Sample condition	Signal	Concentration (ng/mL)	Expected concentration (ng/mL)	%
Dilution linearity	CRC1	18131	8.75	-	-
	CRC1 1:2	15123	7.27	4.37	166
	CRC1 1:3	14984	7.20	3.63	198
	CRC1 1:4	**8480**	**3.96**	**3.60**	**110**
Spike recovery	**CRC1 spike low**	**18702**	**9.03**	**11.85**	**76**
	CRC1 spike high	20201	9.76	14.95	65
	CRC1 1:2 spike low	**16882**	**8.13**	**10.37**	**78**
	CRC1 1:2 spike high	18319	8.84	13.47	66
	CRC1 1:3 spike low	14115	6.77	10.30	66
	CRC1 1:3 spike high	13759	6.59	13.40	49
	CRC1 1:4 spike low	**12626**	**6.03**	**7.06**	**86**
	CRC1 1:4 spike high	**21695**	**10.49**	**10.16**	**103**
Dilution linearity	HV3	972	below LLOD	-	-
	HV3 1:2	762	below LLOD	-	-
	HV3 1:3	786	below LLOD	-	-
	HV3 1:4	736	below LLOD	-	-
Spike recovery	HV3 spike low	2373	0.74	3.10	24
	HV3 spike high	3923	1.59	6.20	26
	HV3 1:2 spike low	3728	1.49	3.10	48
	HV3 1:2 spike high	6766	3.08	6.20	50
	HV3 1:3 spike low	4229	1.76	3.10	57
	HV3 1:3 spike high	7489	3.45	6.20	56
	HV3 1:4 spike low	4292	1.79	3.10	58
	HV3 1:4 spike high	7339	3.37	6.20	54

LLOD: lower limit of detection. CRC1: colorectal cancer sample. HV: healthy volunteer sample.

Given the undetectable levels observed for one of the samples initially tested and recalling what was observed during the ELISA tests described in section 5.4.1, we decided to also test plasma as an alternative matrix for this ECL assay. Besides, taking into account the observation that the 1:4 dilution factor exhibited the best performance in the last test, we also aimed to test a further dilution – 1:8, to assess whether a range of dilution factors with steady linearity is observed. Therefore, in this next experiment (results shown in table 5.18), both plasma and serum from one CRC patient and one individual with Crohn´s disease (CD) were tested. Undiluted and diluted samples (1:4 and 1:8) were analysed, both with and without spiking using the same spike concentrations as before.

Our first observation was that plasma NAP1L1 concentrations were higher than serum levels, as was the case during the ELISA tests. The difference between plasma and serum concentrations was higher for the CRC sample (17-fold) than for the CD sample (1.9-fold). The CD sample was obtained from a bio-bank housed in the Department of Molecular and Clinical Pharmacology at the University of Liverpool and had been stored for a few months, whereas the CRC sample was taken from the Brazilian cohort (storage time: approximately 2 years). This may suggest that the storage time could affect serum levels of NAP1L1 more than plasma levels and explain, at least partially, the plasma/serum difference observed in this study. However, the absence of sample stability tests to assess the effect of storage time and temperature makes it impossible to clarify this point. In either case (plasma levels being truly higher than serum levels or plasma samples being more stable than serum), plasma would be the best matrix for the assessment of this protein. Another result obtained from this experiment was the confirmation that not only the dilution 1:4, but also 1:8 provided adequate linearity and spike recovery for most conditions (highlighted in bold in the table below). Additionally, plasma samples demonstrated better performance than serum samples for quality control parameters, with more conditions exhibiting readings within 75-125% of the expected value.

Table 5.18. Dilution linearity and spike recovery for serum and plasma samples. Results within 75-125% of the expected value are highlighted in bold.

Sample	Signal	Concentration (ng/mL)	Expected concentration (ng/mL)	%
CD004 Plasma	512	0.409	-	-
CD004 P 1:4	**208**	**0.098**	**0.102**	**96**
CD004 P 1:8	**179**	**0.062**	**0.051**	**121**
CD004 P 1:4 low	2488	1.927	3.198	60
CD004 P 1:4 high	**7192**	**4.888**	**6.298**	**78**
CD004 P 1:8 low	**3145**	**2.373**	**3.162**	**75**
CD004 P 1:8 high	**7296**	**4.949**	**6.262**	**79**
CD004 Serum	314	0.216	-	-
CD004 S 1:4	163	0.040	0.054	74
CD004 S 1:8	144	0.009	0.027	35
CD004 S 1:4 low	**3300**	**2.476**	**3.140**	**79**
CD004 S 1:4 high	5779	4.042	6.240	65
CD004 S 1:8 low	2969	2.255	3.109	73
CD004 S 1:8 high	6437	4.440	6.209	71
CR073 Plasma	1142	0.940	-	-
CR073 P 1:4	**350**	**0.253**	**0.235**	**108**
CR073 P 1:8	**249**	**0.146**	**0.118**	**124**
CR073 P 1:4 low	3203	2.411	3.353	72
CR073 P 1:4 high	6947	4.743	6.453	74
CR073 P 1:8 low	**4439**	**3.212**	**3.246**	**99**
CR073 P 1:8 high	**7206**	**4.896**	**6.346**	**77**
CR073 Serum	174	0.055	-	-
CR073 S 1:4	185	0.069	0.014	506
CR073 S 1:8	154	0.026	0.007	383
CR073 S 1:4 low	**3346**	**2.506**	**3.169**	**79**
CR073 S 1:4 high	6110	4.240	6.269	68
CR073 S 1:8 low	**3857**	**2.840**	**3.126**	**91**
CR073 S 1:8 high	**7075**	**4.819**	**6.226**	**77**

The linearity demonstrated for dilutions 1:4 and 1:8 (for plasma samples) established a suitable working range for further tests. The dilution 1:4 was selected due to the low concentrations generally observed, aiming to avoid undetectable readings. Given the linearity and spike recovery demonstrated, plasma was identified as the best matrix for NAP1L1 assessment, a finding concordant with our ELISA results described in section 5.4.1. Therefore, only this matrix was assessed in the next steps of assay development.

5.5.5. **Evaluation of alternative blocking agents**

At this stage of the work, our supply of antibodies had been depleted. We therefore purchased new stocks of both the Abnova mouse monoclonal and the Abcam rabbit polyclonal antibodies. Catalogue numbers and even lot numbers were exactly the same as the ones that we had been using until this point. In order to avoid deviations in assay performance, identical reagents should ideally be employed. Some substances are very sensitive to small variations in production or storage conditions. This is especially true for biological substances such as antibodies. Consequently, some parameters should be retested when such reagents are replaced, even when using similar products from the same manufacturers. We therefore tested the new antibodies using the same assay configuration previously selected. As figure 5.21 shows, an adequate standard curve was again produced, confirming good antigen-antibody reactivity. However, when the readings were carefully analysed (see table 5.19, column related to the 2.5µg/mL concentration), it was noticed that the background signal was considerably higher than that observed when the antibodies that had previously been used. For that reason, strategies aiming to reduce non-specific binding and signal background were needed. Our first attempt was to reduce the capture antibody concentration. Along with the current 2.5µg/mL concentration, we also tested 1.25, 1.0 and 0.5µg/mL. As table 5.19 also shows, these reductions in capture antibody concentration resulted in drops in background signal at the cost of significant decreases in maximum concentration readings, thus reducing the SBR. Consequently, an alternative approach was necessary to reduce signal noise whilst maintaining optimal SRB and dynamic range.

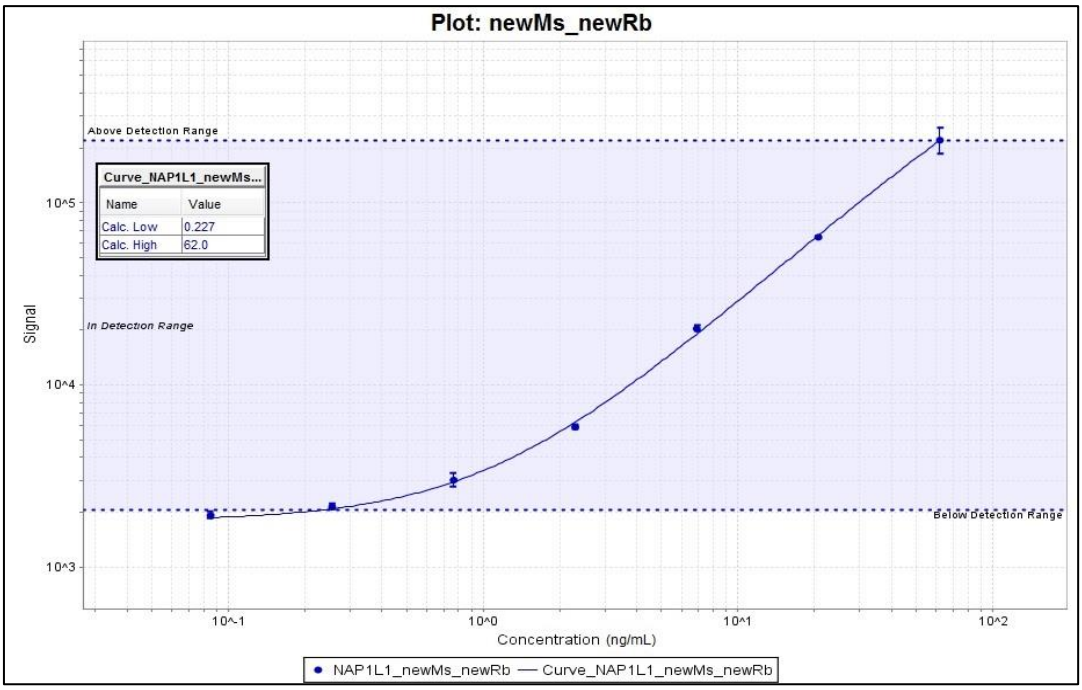


Figure 5.21. Standard curve and dynamic range produced using new vials of capture and detection antibodies. The higher background exhibited by this assay (readings displayed in table 5.20) translated into a compressed dynamic range. The LLOD was 0.23ng/mL, considerably higher than that observed when testing the old stock of antibodies.

Table 5.19. Standard curve readings and SBRs for different capture antibody concentrations. Standard samples tested in duplicate.

		Capture antibody concentration			
		2.5µg/mL	1.25µg/mL	1.0µg/mL	0.5µg/mL
Sample	Concentration	Signal	Signal	Signal	Signal
S001	62.000	195834	50269	24041	9192
S001	62.000	244527	51350	24334	8035
S002	20.667	64938	18686	9007	3453
S002	20.667	64302	15711	8262	3188
S003	6.889	19849	7190	3250	1446
S003	6.889	20837	7249	3111	1435
S004	2.296	5738	2987	1344	777
S004	2.296	6002	3285	1262	905
S005	0.765	2833	1668	745	552
S005	0.765	3172	1732	798	580
S006	0.255	2083	1194	533	499
S006	0.255	2196	1271	625	560
S007	0.085	1865	1113	484	486
S007	0.085	1979	1107	556	438
S008	0.000	1650	1162	412	352
S008	0.000	1793	1041	425	390
	SBR	**127.90**	**46.13**	**57.80**	**23.22**

As previously mentioned, the blocking step performed after capture antibody coating aims to occupy any remaining surface on the plate well with a non-cross-reactant protein, thus preventing nonspecific binding. The standard blocking reagent for MSD assays is the proprietary Blocker A, basically a 1% bovine serum albumin (BSA) solution. Alternatively, we tested 1% casein diluted in either PBS or TBS as a blocking agent. The same solutions were also used for calibrator and sample dilution, as well as detection antibody and tertiary SULFO-TAG™ antibody dilution. Table 5.20 shows the results for this experiment. Both casein blocking solutions resulted in striking reductions (10 to 24-fold) in signal readings across the entire standard curve, confirming strong inhibition of background noise. As this decrease was observed from top to zero concentration, the SBR was again used to identify the best condition. Although casein in PBS resulted in the lowest background, the SBR produced was the highest among the conditions tested.

Table 5.20. Standard curve readings and SBRs for different blocking reagents. Standard samples tested in duplicate.

		Blocking reagent		
		Blocker A	**Casein in PBS**	**Casein in TBS**
Sample	**Concentration**	**Signal**	**Signal**	**Signal**
S001	62.000	195834	19375	7588
S001	62.000	244527	21845	7158
S002	20.667	64938	7095	2474
S002	20.667	64302	7213	2295
S003	6.889	19849	2535	931
S003	6.889	20837	2599	797
S004	2.296	5738	875	353
S004	2.296	6002	910	308
S005	0.765	2833	343	177
S005	0.765	3172	363	180
S006	0.255	2083	174	116
S006	0.255	2196	179	125
S007	0.085	1865	122	109
S007	0.085	1979	120	106
S008	0.000	1650	94	97
S008	0.000	1793	90	101
	SBR	**127.90**	**224.02**	**74.47**

Therefore, one percent casein in PBS overcame the increase in signal background observed with the new stock of antibodies without reducing the dynamic range and, thus, this was selected as the blocking reagent for subsequent experiments.

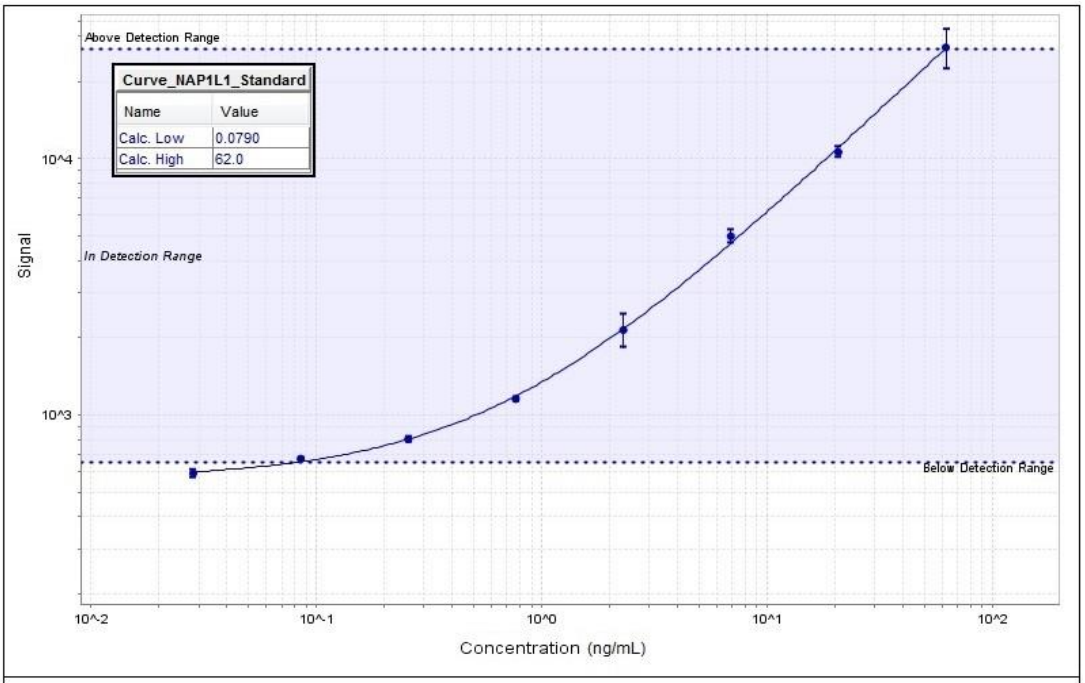
5.5.6. **Final assay configuration and quality control tests**

The previous steps performed in this assay development project allowed us to establish the conditions which resulted in adequate performance in terms of antibody reactivity, sensitivity, low background, high SBR and minimal matrix interference. Figure 5.22 depicts the final standard curve, the limits of detection and the assay conditions. Once a new assay has been developed, it should undergo several quality control tests in order to reassure its performance in the selected matrix and in the relevant clinical scenario (Del Campo *et al*., 2015, Andreasson *et al*., 2015). In this research, the following parameters were assessed:

a. Dynamic range: limits of detection are shown in figure 5.22, below;

b. Precision: the usual measure of precision is the coefficient of variation (CV) for the results observed in repeated experiments or replicates. CV is calculated using the formula CV=(SD/Mean)x100, where SD is the standard deviation. Both intra-assay and inter-assay CVs were calculated;

c. Spike recovery: although already tested in the preliminary phase, recovery was tested again due to the change in the blocking reagent and the replacement of antibodies.



Final assay configuration:

Capture AB: Abnova mouse monoclonal 2.5µg/mL

Detection AB: Abcam Rabbit polyclonal 1.0µg/mL

Calibrator: Abcam human recombinant NAP1L1

Standard curve top concentration: 62ng/mL

Standard curve dilution factor: 3

Sample dilution factor: 4

Blocking agent: Casein 1% in PBS

Reporter: MSD SULFO-TAG™ anti-rabbit AB 1.0µg/mL

Figure 5.22. Final standard curve and assay configuration. The use of 1% casein in PBS along with the same conditions previously established allowed the production of an assay with optimal background, SBR and dynamic range (0.08 – 62.0ng/mL). AB: antibody.

Prior to final spike recovery and inter-assay CV experiments, two plasma samples (one from a normal control and one from a CRC patient) were aliquoted and stored at -80°C so that all aliquots had undergone the same number of freeze-thaw cycles. During six consecutive days, an aliquot from each sample was defrosted and tested with and without spiking. In this

assay, spiking was performed by adding to the diluted sample a volume of calibrator solution (concentration: 31ng/mL) in a 1:1 ratio - the high-concentration spike (HS). Therefore, considering the final solution, the expected concentration was the concentration exhibited by the non-spiked sample divided by 2, plus 15.5ng/mL. Additionally, we explored a low concentration spiking (LS), repeating the procedures described above using a calibrator solution at 7.5ng/mL. The expected concentration, in this case, was the sample concentration divided by 2, plus 3.75ng/mL. The CRC sample was tested under both spiking conditions, whereas the normal sample underwent the HS analysis only due to the scarce amount of normal plasma available at this late stage of development. It has been determined that recovery ranges from 70-80% to 120-130% were acceptable (Leary *et al*., 2013). As table 5.21 demonstrates, spike recovery for the conditions tested was between 111% and 115%, thus fulfilling this quality control criterion. Inter-assay CV expresses plate-to-plate consistency and should ideally be < 20% (Leary *et al*., 2013). It was calculated as the mean of the CVs observed for each sample condition in the six daily experiments. Mean inter-assay CV was 13.2% in this series of tests. It also satisfies the accepted criterion for imprecision and matches the performance disclosed by commercially available NAP1L1 ELISA kits (CV = 12%).

Table 5.21. Final tests of spike recovery and inter-assay CVs.

	DAY 1	**DAY 2**	**DAY 3**	**DAY 4**	**DAY 5**	**DAY 6**	**Mean**	**SD**	**CV (%)**
N 1:4	5.20	5.95	5.88	6.48	4.31	4.26	**5.35**	0.92	**17.16**
N 1:4 HS	19.50	20.10	20.40	21.50	17.80	18.70	**19.67**	1.31	**6.64**
(Recovery)	111%	112%	114%	118%	104%	109%	**111%**	4.71	**-**
CRC 1:4	11.30	16.30	16.50	18.00	11.50	11.20	**14.13**	3.12	**22.11**
CRC 1:4 HS	25.90	26.70	25.60	27.00	22.60	24.30	**25.35**	1.65	**6.50**
(Recovery)	125%	115%	110%	113%	109%	118%	**115%**	6.08	**-**
CRC 1:4 LS	12.50	14.80	12.80	11.50	10.30	10.80	**12.12**	1.63	**13.42**
(Recovery)	133%	124%	107%	90%	108%	116%	**112%**	14.94	**-**
								Mean CV:	**13.17**

HS: high-concentration spike. LS: low-concentration spike.

Additionally, intra-assay CV was calculated comparing the variation between the replicates of the 80 plasma samples tested in the final assays. The accepted range of variation for this parameter is < 10%. A mean intra-assay CV of 6.65% was obtained, thus, again, showing an adequate performance.

Some important validation tests were not performed in this research due to time constraints. While repeatability – measurements performed under the same conditions, was reassured by the daily tests, reproducibility – measurements under different conditions (Rifai *et al*., 2006), was not fully tested. Sample storage stability was also not fully assessed, preventing us from knowing what is the effect of length and temperature of storage, or the number of freeze-thaw cycles on the measured level of NAP1L1. Finally, although dilution linearity and parallelism were assessed in the preliminary phase, it was limited to a maximum of 8-fold dilution. Testing further dilutions could establish a wider working range and provide a better view of the assay performance in the target matrix. Despite these issues, the overall consistency and accuracy demonstrated by this novel immunoassay outperforms the ELISA kits described in previous sections, especially in terms of sensitivity – exactly the major flaw exhibited by those kits. Therefore, we deemed this assay adequate for sample typing and decided to proceed with the testing of clinical samples.

5.5.7. **Assessment of NAP1L1 concentrations in clinical samples**

Using the conditions depicted in figure 5.22, we assessed NAP1L1 plasma concentrations in our clinical samples. For this analysis, we used the same samples tested using the commercial ELISA kits. Therefore, for the Brazilian cohort, we tested normal controls (n=10) and cancer samples (n=30) as these were the most representative groups in that sample set. For the UK cohort, we used samples from four clinical groups: normal controls (n=10), low-grade adenomas (n=10), high-grade adenomas (n=7) and cancers (n=13).

Results from the Brazilian cohort are illustrated in figure 5.23. No significant difference in NAP1L1 concentration was observed between normal control and cancer groups. A remarkably high variation in results was observed among samples within clinical groups. Mean protein concentration (and 95% confidence interval) were 51.5ng/mL (26.2 – 129.3ng/mL) for normal controls, and 69.4ng/mL (33.6 – 105.2ng/mL) for cancer samples. Minimum and maximum concentrations were 1.23 and 357.5ng/mL in the normal group, and 0.4 and 314.8ng/mL in the cancer group.

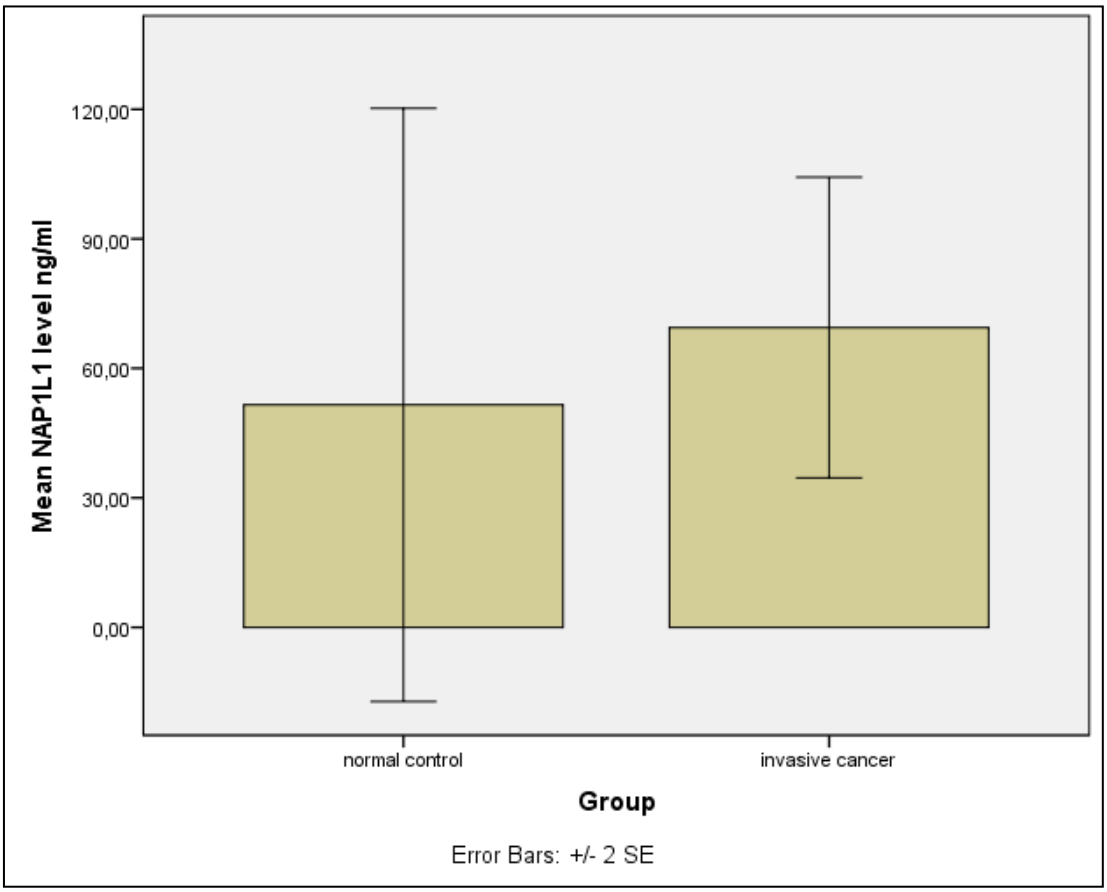


Figure 5.23. Mean NAP1L1 concentrations in the Brazilian cohort. A large variation in protein levels was observed, as depicted by the error bars (+/- 2 standard errors). No difference in protein concentration was found when comparing the groups (p = 0.44, Mann-Whitney U test).

Similarly to the Brazilian cohort, the results for the UK cohort also exhibited large intra-group variation, as shown in figure 5.24. Mean NAP1L1 concentration (and 95%CIs) for the groups tested were 184.0ng/mL (63.4 – 304.5ng/mL) for normal controls, 175.4ng/mL (80.1 – 270.6ng/mL) for low-

grade adenomas, 283.1ng/mL (89.1 – 477.0ng/mL) for high-grade adenomas and 305.9ng/mL (200.2 – 411.6ng/mL) for the cancer group. A clear trend of increased concentration in high-grade adenomas and cancer samples was observed. However, this difference was not statistically significant (p = 0.21, one-way ANOVA). This result is strikingly similar to the observations obtained when testing the UK plasma samples using the DL Develop and the Abbexa NAP1L1 ELISA kits (compare the figure below with figures 5.3 and 5.6 in section 5.4.1), although the magnitude of the levels is significantly higher here.

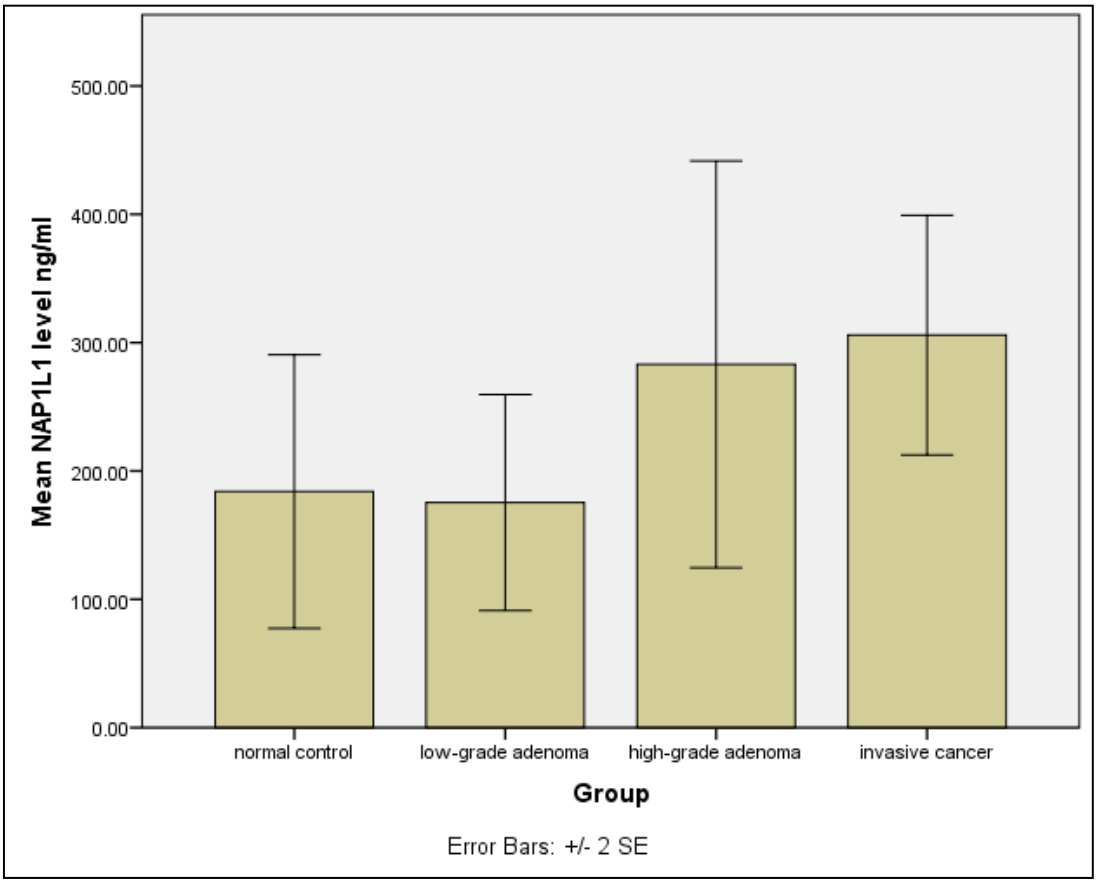


Figure 5.24. Mean NAP1L1 concentrations in the UK cohort. A trend of increased protein concentration was observed, although this was not statistically significant (p = 0.21, one-way ANOVA). Again, protein concentration exhibited large variation within groups.

Additionally, NAP1L1 levels were considerably higher in the UK cohort when compared to the Brazilian samples, a finding already observed during the ELISA experiments. Given the low inter-assay CV (making it possible to

compare the results from different plates), it reinforces the hypothesis that storage length and conditions may affect NAP1L1 readings in immunoassays, as these two sample cohorts were collected at different periods and locations.

This research, through the experiments described in this sub-chapter, resulted in the development of a novel ECL assay for NAP1L1 based on the MSD platform that successfully fulfills the major criteria for immunoassay validation. However, it failed to demonstrate a significant difference in protein expression between the clinical groups tested. Although a trend towards increased NAP1L1 concentration in patients with high grade dysplastic and malignant lesions was consistently observed, the heterogeneity of results impeded statistical confirmation of differential expression. The accuracy, consistency and sensitivity exhibited by this novel ECL assay reassure the robustness of these results. It suggests that the pursuit of alternative techniques is unlikely to provide different results if the same sample types are assessed. An unanswered question is whether the assessment of fresh samples or samples that had been stored for shorter periods of time would yield different results. Further research is, therefore, recommended to clarify this issue.

5.6. Discussion

This chapter describes experiments that were performed to assess whether NAP1L1, RPL6 and PHB are potential CRC blood biomarkers. The analysis of two cohorts from different locations was performed in an attempt to obtain robust and valid results. The initially selected platform was the ELISA, a widely available method for the assessment of protein concentrations in biological fluids (Lequin, 2005, Rifai *et al*., 2006). As no validated kit for the assessment of our candidate biomarkers had been reported, our initial task was to search for suitable options among the commercially available kits. Our starting point was the same plate that had been tested previously by our group. However, issues related to the

consistency and reproducibility of those kits demanded the exploration of different options in the form of both alternative ELISA kits and other immunoassay methods.

RPL6 and PHB ELISA kits demonstrated good quality in terms of accuracy and consistency. However, for both proteins, initial experiments using serum samples demonstrated insufficient sensitivity, as several results were below the detection range. The use of plasma instead of serum resulted in higher protein concentrations for both markers using different ELISA kits, thus suggesting that plasma would be the favoured matrix for their assessment. When using plasma, valid results were obtained for most samples, thus allowing the comparison of protein levels in these sample cohorts. For the Brazilian cohort, normal controls and cancer samples were compared, whilst for the UK cohort samples from four clinical groups – normal controls, low- and high-grade adenomas and cancers, were analysed. However, the assessment of the protein content did not show any significant differences between those clinical groups. These results are in disagreement with the findings previously demonstrated by our research group. The most probable cause for this difference is the sample sizes evaluated. In the preliminary research described at the beginning of this chapter, only two cancer cases (indeed, one cancer patient and one patient with a previous CRC and a possible recurrence) were assessed along with 37 non-cancer affected individuals who had various other health conditions. In the present work, more cancer samples (30 from the Brazilian cohort and 13 from the UK cohort) were analysed. Pre-neoplastic lesions (high-grade adenomas) and low risk-lesions (low-grade adenomas) were also assessed, but no other disease state was included. Therefore, this analysis is much more empowered and trustworthy than the preliminary work and is probably a better reflection of reality. Given the consistency of results exhibited by plasma samples, we accepted the null hypothesis in these cases, and concluded that both RPL6 and PHB are not differentially expressed in blood from individuals with CRC or adenomas when compared with normal individuals.

However, possible interference caused by differences in sample collection and storage cannot be completely ruled out. For both markers, the protein levels in the UK cohort were higher than in the Brazilian cohort. Both sample sets were collected and stored according to the same protocols. However, the collections were performed in different locations, mostly by local staff, and minor but potentially relevant differences in procedures may have occurred. In addition, the major difference between the two cohorts was the length of sample storage, 1-3 years for the Brazilian cohort compared to weeks to a few months for the UK cohort (even less for the samples used in the preliminary phase). Important differences in protein reactivity may ensue as a result of diverse sample storage length (Macaraeg *et al*., 2015). As, due to time constraints, we were not able to perform sample stability tests, the effect of the length and the temperature of storage on sample results could not be ascertained. Therefore, the assessment of fresh samples or the analysis of sample stability would be recommended to resolve this question.

All three NAP1L1 ELISA kits tested in this thesis (from Cloud Clone Corp, DL Develop and Abbexa) exhibited poor performances. Low consistency, inadequate sensitivity, and poor correlation were generally observed. The only consistent finding from these experiments was the observation that plasma NAP1L1 concentrations were always higher than serum concentrations, as was the case for RPL6 and PHB. Due to the poor quality demonstrated by these kits, we decided to develop our own in-house immunoassay to properly test NAP1L1 concentration in our clinical samples. For this purpose, we selected the MSD-ECL platform. This method has been used for assay development by several research groups, with results generally superior to ELISAs (Chaturvedi *et al*., 2015, Sloan *et al*., 2012, Postelnek *et al*., 2016). Immunoassay development protocols follow a step-wise approach to test each reagent used under different conditions to obtain an acceptable performance. In this research, we tested several commercially available antibodies as capture and detection reagents. The antibody pair of Abnova mouse monoclonal - Abcam rabbit polyclonal demonstrated a good affinity for the calibrator (human recombinant NAP1L1 from Abcam), and resulted in adequate assay linearity for the standard curve. Using all the

optimisation procedures and conditions described in the previous sections, an MSD-ECL assay for NAP1L1 was successfully developed. It passed major quality control tests, with an overall performance surpassing the ELISA quality control parameters demonstrated in our study. In particular, the improved sensitivity and lower limit of detection obtained allowed the quantification of the protein content in all plasma samples tested.

After confirming that this assay was an adequate method for protein quantification, final analysis of the clinical samples was performed. The same samples were tested as those that had been tested using the ELISA kits. The final assessment using this novel in-house immunoassay did not show a significant difference in NAP1L1 concentration between normal and cancer samples in the Brazilian cohort. For the UK cohort, similarly, no significant difference was observed when comparing the four groups tested. A trend towards increased plasma NAP1L1 concentrations in high-grade adenomas and cancer samples was observed, but the prominent heterogeneity between results within groups prevented statistical confirmation. Notably, this was exactly the same result as that obtained during the NAP1L1 ELISA tests using two different plates (DL Develop and Abbexa). Some of the other findings made during the ELISA experiments were also confirmed. Notably, NAP1L1 concentration was higher in plasma than serum in the ECL experiments. It is, therefore, highly recommended that any future research assessing NAP1L1 concentrations in the blood uses plasma as the preferred matrix for testing this protein. Additionally, the UK samples exhibited higher protein levels than the Brazilian samples. As discussed above, this may be due to differences in sample collection and storage. Stability tests are suggested in order to clarify this point.

The experiments described in this chapter have therefore not succeeded in identifying any of these candidate proteins as CRC blood biomarkers. However, the results have provided interesting insights for the assessment of NAP1L1, RPL6 and PHB concentrations in the blood. The knowledge generated by this research may help future studies that plan to measure the concentrations of these proteins in biological fluids. It also highlights the superior performance of ECL-based methods, particularly using

the tools provided by MSD, in the development of customised immunoassays. Given the large variation in results, the investigation of these proteins, especially NAP1L1, as a prognostic rather than a diagnostic marker could result in clinically relevant findings. Unfortunately, our sample cohorts were very limited in terms of size and time of follow-up after diagnosis so would not permit such investigation at this time.

Chapter Six:
Prognostic significance of the IHC expression of the biomarkers in CRC

6. CHAPTER 6 – PROGNOSTIC SIGNIFICANCE OF THE IHC EXPRESSION OF THE BIOMARKERS IN CRC

6.1. Introduction

As discussed in *Chapter 1*, a biomarker is "a characteristic used to measure and evaluate objectively normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention" (Atkinson *et al*., 2001). Prognostic biomarkers are clinical or laboratorial characteristics that are capable of predicting clinical outcome (Pritzker, 2015, Atkinson *et al*., 2001). In oncology, the most relevant clinical outcome is overall survival (Sherrill *et al*., 2012), although there are other important endpoints such as tumour recurrence, progression free survival, the need for specific medical interventions and quality of life. CRC prognostic biomarkers have been extensively researched and a multitude of candidates have been suggested in the literature. However, very few of these have been incorporated into routine clinical practice (Bianchi *et al*., 2011, Watson and Søreide, 2016, Sagaert, 2014). A major marker of Wnt pathway activation, β-catenin nuclear immunostaining has been associated with worse prognosis is several studies of CRC, as confirmed by a large meta-analysis (Chen *et al*., 2013). Therefore, we hypothesised that our candidate biomarkers (which are derived from Wnt activation models) might also be related to clinical outcome in CRC patients.

In *Chapter 3*, we demonstrated that NAP1L1, RPL6 and PHB exhibited differential patterns of expression between malignant and non-malignant colorectal tissues. Nonetheless, the intensity of protein expression within each sample group, particularly in malignant tissues, exhibited wide variations, thus highlighting the heterogeneity of biomarker expression. Similar findings were observed in the assessment of both the tissue transcriptomic and blood proteomic expression of these candidate markers (*Chapters 4* and *5*). If the expression of these genes is related to malignant characteristics (as suggested by the mechanistic studies also described in *Chapter 4*), it is intuitive to consider that this heterogeneous expression might

result in different biological behaviour. Thus, the content and/or localisation of these proteins in tumour tissues might be potentially useful prognostic markers.

Proteomic methods are promising strategies for biomarker discovery and validation (Li and Chan, 2014, Tjalsma, 2010, Alvarez-Chaver *et al*., 2014, de Wit *et al*., 2013). Amongst the several methods for assessing protein expression, IHC is relatively inexpensive, reproducible, and widely available (Lin and Chen, 2014). Other advantages and issues related to IHC have been described in *Chapter 3*. Despite some pitfalls, the method is currently used for the assessment of diagnostic, prognostic and predictive biomarkers in several types of cancers (Chamberlain *et al*., 2015, Zaha, 2014, Toffart *et al*., 2014, Varma and Jasani, 2005).

We therefore performed an assessment of the relationship between the IHC expression of our candidate proteins and clinicopathologic variables (age, gender, tumour grade and stage). Additionally, we also analysed whether the expression of these biomarkers correlated with overall survival in CRC patients.

6.2. Samples and immunohistochemistry procedures

For this assessment, the main inclusion criteria were: 1. diagnosis of CRC and; 2. four or more years of follow-up. This time frame was selected because most CRC recurrences occur within 2 to 3 years of follow-up (Buie and Attard, 2005). Exclusion criteria were: 1) death occurring within 30 days of surgery, as this variable identifies deaths that are related to surgical complications, not to the cancer (Jacobs *et al*., 2006); 2) administration of chemotherapy or radiotherapy prior to surgery; 3) absence of tumour tissue in the paraffin block. We therefore initially retrieved 95 paraffin blocks from two pathology labs in Cuiaba/Brazil. Patients had been diagnosed with CRC between 2004 and 2012. By using the selection criteria above, we excluded 20 of these cases mainly due to early death or absence of tumour tissue. Consequently, the final prognostic cohort was made up of 75 CRC cases.

Median follow-up time was 84.7 months (ranging from 48 to 153 months). Information regarding pathological diagnosis (date of diagnosis, stage and grade) and clinical characteristics (age and gender) were obtained from the original pathology report and from the records of the pathology labs. The main clinical event studied in this analysis was overall survival. Whenever possible, records from the cancer clinics where the patients had been followed up were assessed in order to obtain additional information about mortality. This strategy was complemented by assessing the Mortality Information System ("*Sistema de Informação de Mortalidade*" – *SIM*), the Brazilian nationwide electronic death registry. A WHO report published in 2005 has classified this database as of "medium quality" (70-90% coverage), ranked in the same category as several high-income European countries (Mathers *et al*., 2005). A more recent report has shown that this system has improved considerably since then, and now covers more than 95% of deaths (Figueiroa Bde *et al*., 2013). Therefore, we assume that the use of this database to confirm the occurrence of death makes our analysis robust. Overall survival was recorded as the interval between diagnosis and death from any cause or the date when the registry was last checked (when death had not occurred). Given the relatively small number of cases included, cancer stages were grouped into two groups: early-stage (stages I and II) and late-stage (stages III and IV). Table 6.1 describes the clinicopathologic characteristics of the patients included in this analysis.

Table 6.1. Characteristics of the patients included in the prognostic analysis.

Characteristics	**Patients (n=75)**
Mean age (range)	59.7 (33-84)
Gender	
Male	43 (57.3%)
Female	32 (42.7%)
Stage	
I-II	28 (37.3%)
III-IV	47 (62.7%)
Grade	.
well differentiated	23 (30.6%)
moderately differentiated	50 (66.6%)
poorly differentiated	2 (2.6%)

The same antibodies used in the initial IHC assessment of the candidate biomarkers were employed in this aspect of the research. The same antibody concentrations were also used. However, the procedures performed in this section were carried out according to the routine practices of the Sao Nicolau pathology lab - Cuiaba/Brazil (*Chapter 2*). Most clinical pathology labs currently use pre-optimised solutions for the various steps involved in IHC staining in order to save time and increase productivity. Therefore, it would be interesting to assess the validity of our candidate biomarkers in this "real-world" setting. For this purpose, we first assessed the pattern of staining produced by this technique compared with the pattern obtained during our initial IHC study (*Chapter 3*) using a few samples. Fortunately, the staining pattern observed was very similar using both methods for the majority of the candidate proteins tested, thus permitting us to continue with this method assessment. Differences in image background and colour shade (as a result of the use of different cameras in different time points) were compensated by the "*threshold setting"* and the "*background correction*" functions of the scoring plugins. These functions permit adjustments that counterbalance differences caused by different equipment being used for photographing the slides (Tuominen *et al*., 2010).

For assessment of the stained IHC slides, we again used the electronic scoring tools incorporated in ImageJ: IHC Profiler for cytoplasmic staining and ImmunoRatio for nuclear staining. Both methods have been thoroughly explained in *Chapter 3*. Examples of nuclear and cytoplasmic scores are depicted in figure 6.1. After scoring the slides, Receiver Operating Characteristics (ROC) curves with nuclear and cytoplasmic results were plotted for each marker. Scores exhibiting the best trade-off between sensitivity and specificity considering the binary endpoint mortality (dead/alive) were identified. These cut-offs were used for classifying samples as having low or high protein expression. Once samples had been split into these two groups, correlations between biomarker expression and clinicopathologic variables were assessed. Furthermore, a survival analysis was also performed to assess whether the expression of these proteins could predict survival.

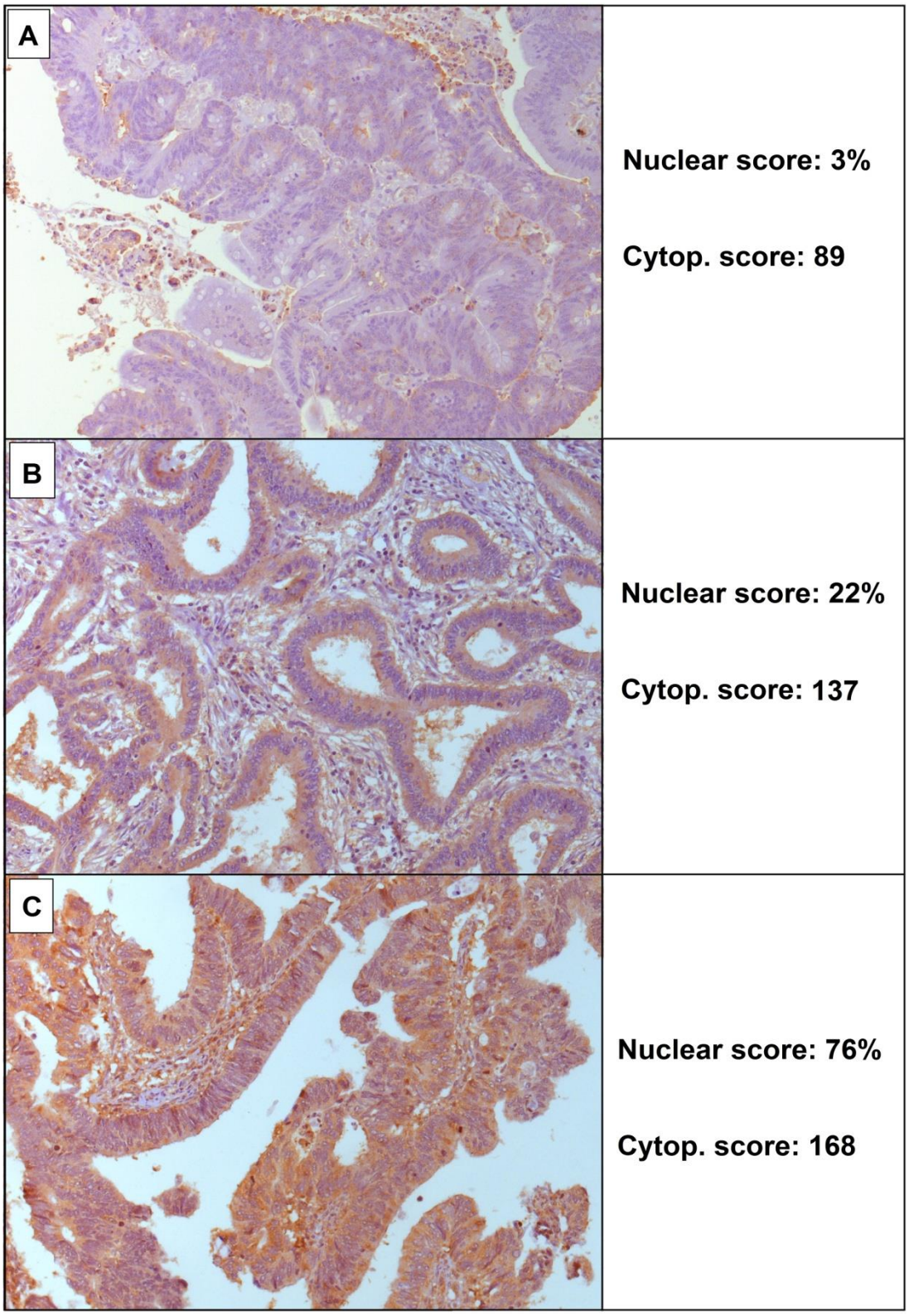


Figure 6.1. Examples of scoring results produced by IHC Profiler and ImmunoRatio in NAP1L1-stained colorectal tissues. The electronic tools were capable of discriminating samples with different patterns of staining. (A) low nuclear and low cytoplasmic staining; (B) low nuclear and high cytoplasmic staining; (C) high nuclear and high cytoplasmic staining (see cut-off values below). Magnification: 200x.

6.3. NAP1L1 as a prognostic marker in CRC

The immuno-expression of NAP1L1 was similar using both staining protocols tested in this research project (figure 6.2). Protein content was scored using IHC Profiler and ImmunoRatio plugins for cytoplasmic and nuclear staining, respectively. Initially, using mortality status as the binary event of interest, ROC curves were plotted. The area under the curve (AUC), representing the accuracy of the test for predicting the status, was 0.58 for the nuclear score and 0.60 for the cytoplasmic score, denoting that these two variables were not very accurate tools for discriminating mortality status. This is probably the result of many patients still being alive by the end of the study in both the low and high expression groups (see figures below). However, the assessment of the ROC curves permitted the selection of a cut-off value with adequate balance between sensitivity and specificity. The NAP1L1 IHC cut-off for nuclear staining was 32% (of positive nuclei). The use of this threshold resulted in a sensitivity of 61% and a specificity of 67.5% for discriminating mortality status. For cytoplasmic staining, a cut-off of 135 (in a range from 0 to 300) yielded a sensitivity of 57% and a specificity of 54%. As the primary aim of this study was predicting survival as measured by the time from diagnosis until death or the end of follow-up and not solely the binary mortality status, these parameters were deemed adequate.

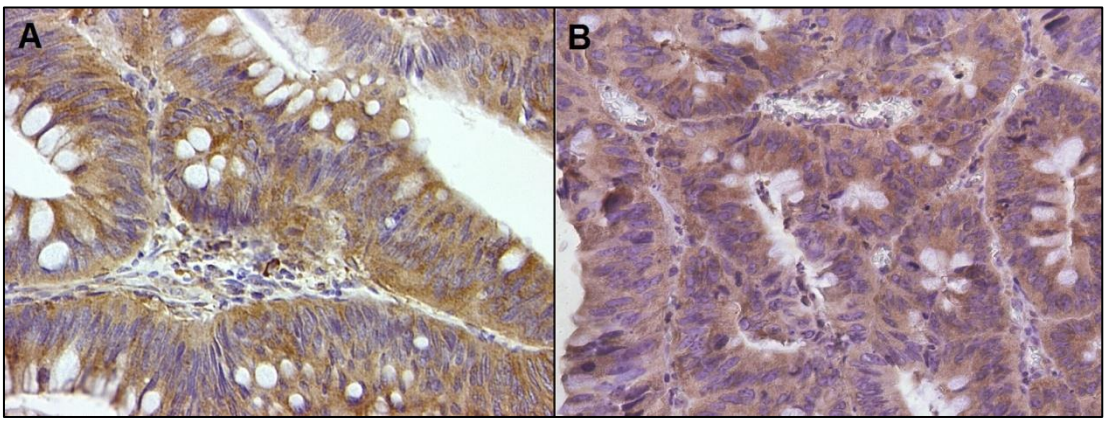


Figure 6.2. Comparison of NAP1L1 staining patterns observed with the IHC protocols used in the initial validation study (A) and in the prognostic study (B). Similar staining intensity and localisation are observed. Differences in colour shade and background are due to the use of different cameras for recording the images. Magnification: 600x.

For the evaluation of nuclear staining, the use of a 32% cut-off resulted in the separation of the sample cohort into low and high-expression groups. In order to assess whether the resulting groups were well balanced in terms of known prognostic factors, we produced a table displaying the main variables (table 6.2) and statistical tests were performed. As can be noted, no significant difference in the distribution of clinicopathologic variables was observed, indicating that no associations exist between NAP1L1 nuclear expression and the variables age, gender, grade and stage. Therefore, any prognostic association eventually found could only be attributed to the biomarker expression pattern. Similar analyses were also performed for the cytoplasmic scoring results and, once more, no significant differences were found (data not shown to avoid unnecessary repetition).

Table 6.2. Clinicopathologic characteristics according to NAP1L1 nuclear expression.

Characteristics	**Low nuclear expression** (n=34)	**High nuclear expression** (n=41)	**(2-sided p values)**
Mean age	61.9	57.8	0.157
Gender			0.648
Male	17	24	
Female	16	18	
Stage			0.338
I-II	10	18	
III-IV	23	24	
Grade			0.351
well differentiated	9	14	
moderately differentiated	24	26	
poorly differentiated	0	2	

No significant difference between groups was observed. Mean age was compared by t-test. Categorical variables were compared by Chi-square test or Fisher´s exact test.

Using the Kaplan-Meier method, cumulative survivals for the two groups (high and low nuclear expression) were compared. Initially, groups

were assessed as a whole, i.e. combining all cases regardless of disease stage. Results are depicted in figure 6.3. A clear difference in cumulative survival was observed when analysing nuclear NAP1L1 staining (p=0.012, log-rank test). In the multivariate analysis including age, gender, stage and grade (Cox proportional hazards model), the nuclear score was independently associated with cumulative survival. The high nuclear expression group exhibited a hazard ratio (HR): 0.39 ([95%CI: 0.17 – 0.87]; p=0.02), denoting a 61% reduction in cumulative mortality in this group. As a result, the estimated 5-year survival was 44.4% in the low expression group and 75% in the high expression group. Median duration of survival was 32 months in the low expression group, whilst it has not been reached for the high expression cohort. The only additional variable also associated with survival was tumour stage (HR: 2.55 [95%CI: 1.01 – 6.43]; p=0.047), an expected finding since stage is a known prognostic factor in CRC. These findings strongly suggest an association between NAP1L1 nuclear staining and survival in CRC patients.

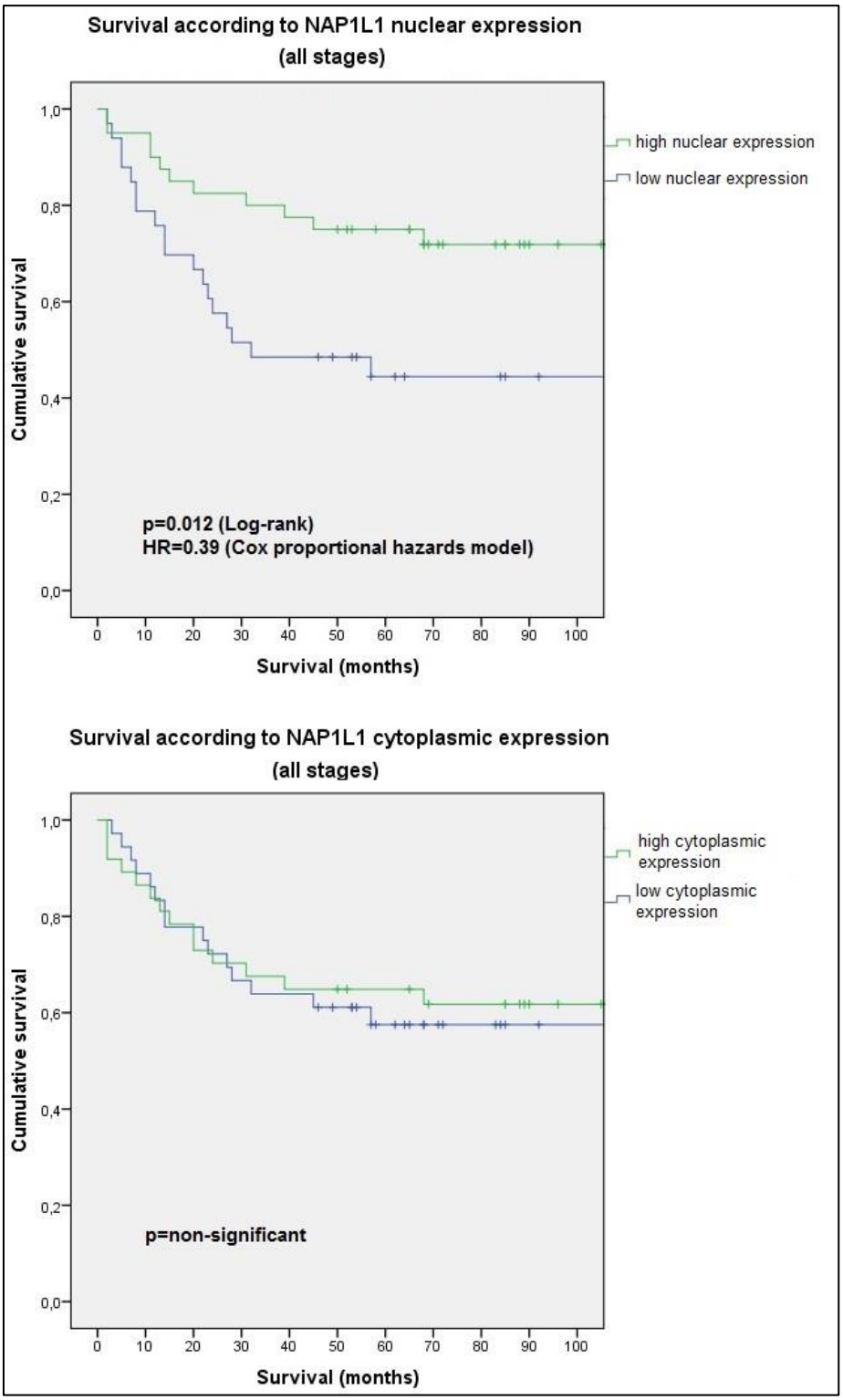


Figure 6.3. Cumulative survival according to NAP1L1 expression (all stages combined). In the superior graph, nuclear staining is assessed. A highly significant (p=0.012) and clinically relevant (HR=0.39 [95%CI: 0.17 – 0.87]) difference in survival between groups was observed favouring the high expression group. In terms of cytoplasmic staining (inferior graph), no significant difference was observed between the low and high expression groups. Vertical marks=censored cases. HR: hazard ratio. CI: confidence interval.

Cytoplasmic NAP1L1 staining was not associated with survival or any other clinicopathologic variable. Kaplan-Meier curves exhibited similar shapes and actually crossed over each other (p=0.759, log-rank test). No further statistical tests were performed given these negative results.

Although a positive association between NAP1L1 nuclear staining and survival was already evident from this initial analysis, we recognise that the patient groups encompassed a quite heterogeneous population in terms of tumour stage. Patients with cancers from stage I (tumours limited to the inner portions of the intestinal wall) to stage IV (metastatic disease) were assessed together. Given the prominent differences in survival between these different stage groups (see figures 1.15 and 1.16 in *Chapter 1*), we decided to analyse patients with early stage (stages I and II) and late stage (stages III and IV) disease separately. This strategy could validate NAP1L1 as a prognostic biomarker for all cancer stages or identify a subgroup more likely to benefit from its application.

In figure 6.4, the results of the analysis of NAP1L1 nuclear staining in either early or late stage tumour groups are shown. For early stage disease, no significant difference in survival was noticed when comparing low and high expression groups (top graph). Conversely, a highly significant (p=0.012, log-rank) difference in survival was observed for the groups involving stages III and IV tumours. Multivariate analysis once again demonstrated that NAP1L1 nuclear score was an independent prognostic factor in CRC patients. The calculated HR (0.28 [95%CI: 0.11 – 0.71]; p=0.008) was even more notable than that observed for the entire cohort, now suggesting a 72% reduction in cumulative mortality. The 5-year survival advantage for high expression tumours was also greater: 70%, versus 34% for low expression cancers. Median survival was only 23 months in the low expression group and, again, has not been reached in the high expression cohort.

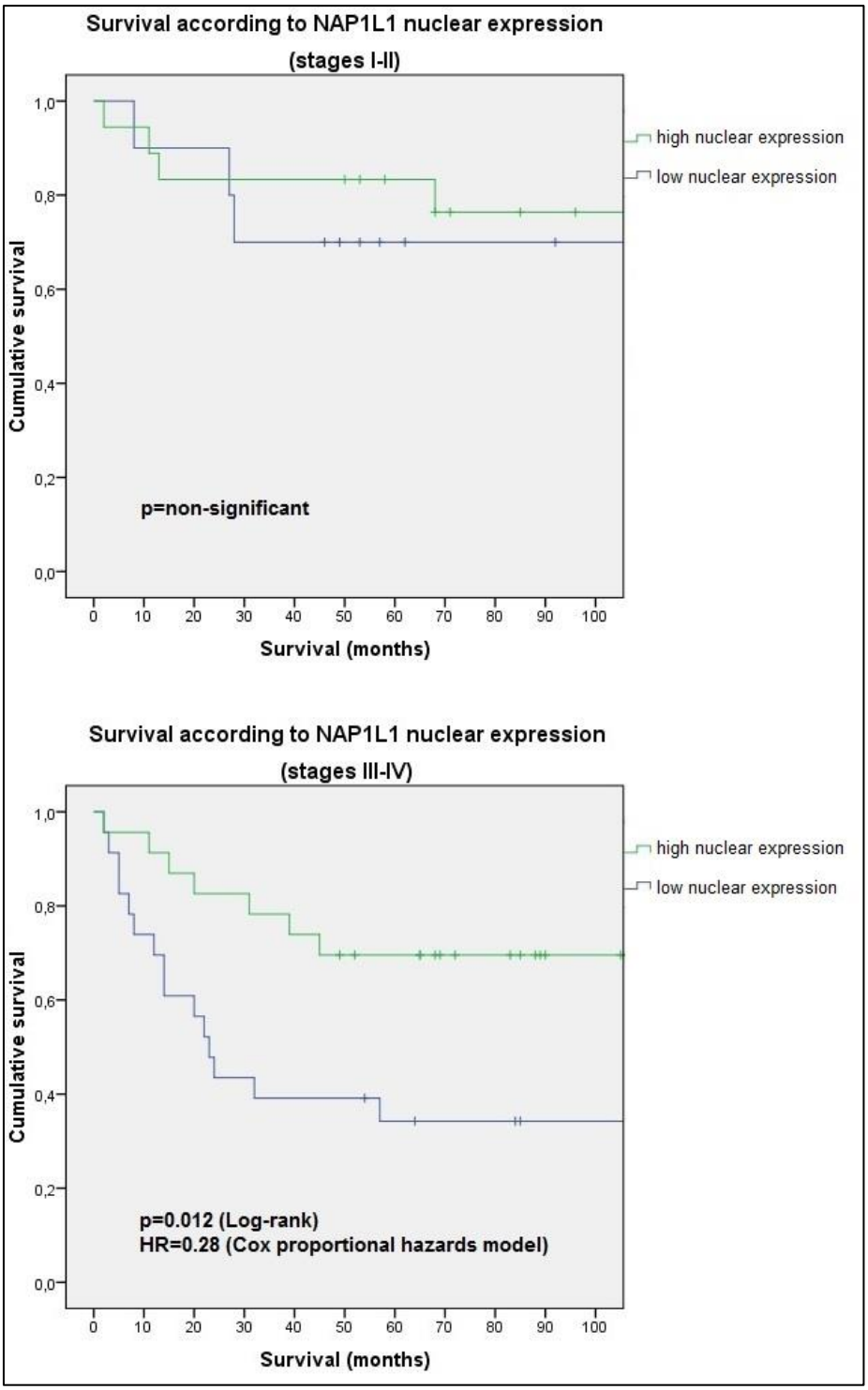


Figure 6.4. Cumulative survival according to nuclear NAP1L1 staining in different stage groups. Top graph shows the results for early stage disease (stages I and II). No difference in survival between low and high expression groups was observed. Bottom graph shows the results for late stage disease (stages III and IV). A statistically significant difference was observed favouring the high expression group (HR: 0.28 [95%CI: 0.11 – 0.71]). Vertical marks=censored cases. HR: hazard ratio. CI: confidence interval.

These findings are not surprising. As the prognosis of late stage cancers is intrinsically worse, more deaths are expected to occur in this group. Survival analyses such as the Kaplan-Meier method are largely dependent upon the occurrence of the event of interest (death, in this case). In fact, the power of the analysis is more related to the number of events than the number of individuals participating in the study (Bradburn *et al*., 2003). Consequently, subgroups with more events are more likely to produce different survival curves when a prognostic association really exists. A longer follow up or a much larger cohort might increase the number of events and identify a prognostic association even in early stage disease. However, our results do not allow us to make such an assumption. Nonetheless, the role of nuclear NAP1L1 staining as a prognostic CRC biomarker, at least for late stage disease, has been clearly demonstrated in a robust and adequate manner. The uncomplicated nature and wide availability of the technique employed (IHC) makes this novel marker an attractive tool for assessing prognosis in CRC patients in clinical settings. However, validation studies to test the reproducibility of these findings in independent cohorts of patients are required in the first instance.

6.4. RPL6 as a prognostic marker in CRC

The same procedures described for NAP1L1 were adopted for the assessment of RPL6 as a potential prognostic marker. A comparison of the staining pattern obtained using the IHC techniques adopted in the Sao Nicolau lab with the staining observed in the initial validation phase of this study indicated different results to those expected. The Sao Nicolau IHC demonstrated a weaker staining pattern than that observed in the validation study (example in figure 6.5). The use of a higher antibody concentration (1:100) did not improve the results. The samples used in this initial comparison were collected approximately 3 years before performing this experiment. To have a better assessment of the staining pattern produced using these conditions, we performed IHC reactions in the entire prognostic

cohort. Surprisingly, almost all (except three) of the 75 samples tested exhibited totally negative RPL6 staining. Even the fragments of normal adjacent mucosa (a type of tissue universally positive in the initial assessment, *Chapter 3*) showed no protein content. We considered the possibility that the commercial solutions used in this experiment were not compatible with the RPL6 antibody tested. Another possible explanation was the different lengths of storage of the FFPE blocks. The sample set used in the initial IHC validation was 1-3 years old, whilst the samples used in this prognostic study had been stored for 4 to 12 years. To test both hypotheses, we re-tested six samples from the prognostic cohort along with six relatively new samples from the initial validation study using the protocol planned for the prognostic study. All cases from the prognostic cohort (between 10-12 years of storage) were negative for RPL6, whilst all cases from the initial sample set (2 years of storage) demonstrated various degrees of positive results (examples depicted in figure 6.6). Notably, samples with fragments of adjacent mucosa (that could be regarded as internal "positive controls" based on our initial results) also demonstrated the same patterns, thus confirming that the absence of immunostaining was limited to the oldest samples.

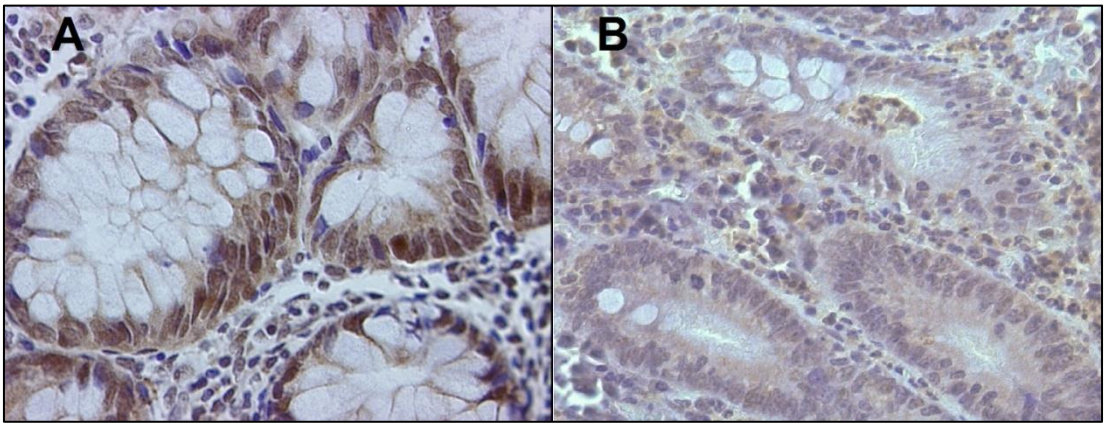


Figure 6.5. Comparison of RPL6 staining patterns observed with the IHC protocols used in the initial validation study (A) and in the prognostic study (B). Weaker staining was observed in the experiment performed in the prognostic cohort. An increase in antibody concentration did not produce different results (data not shown). Noteworthy, these two staining procedures were performed two years apart. Therefore, the sample had been stored for 1 year in A and for 3 years in B. Magnification: 600x.

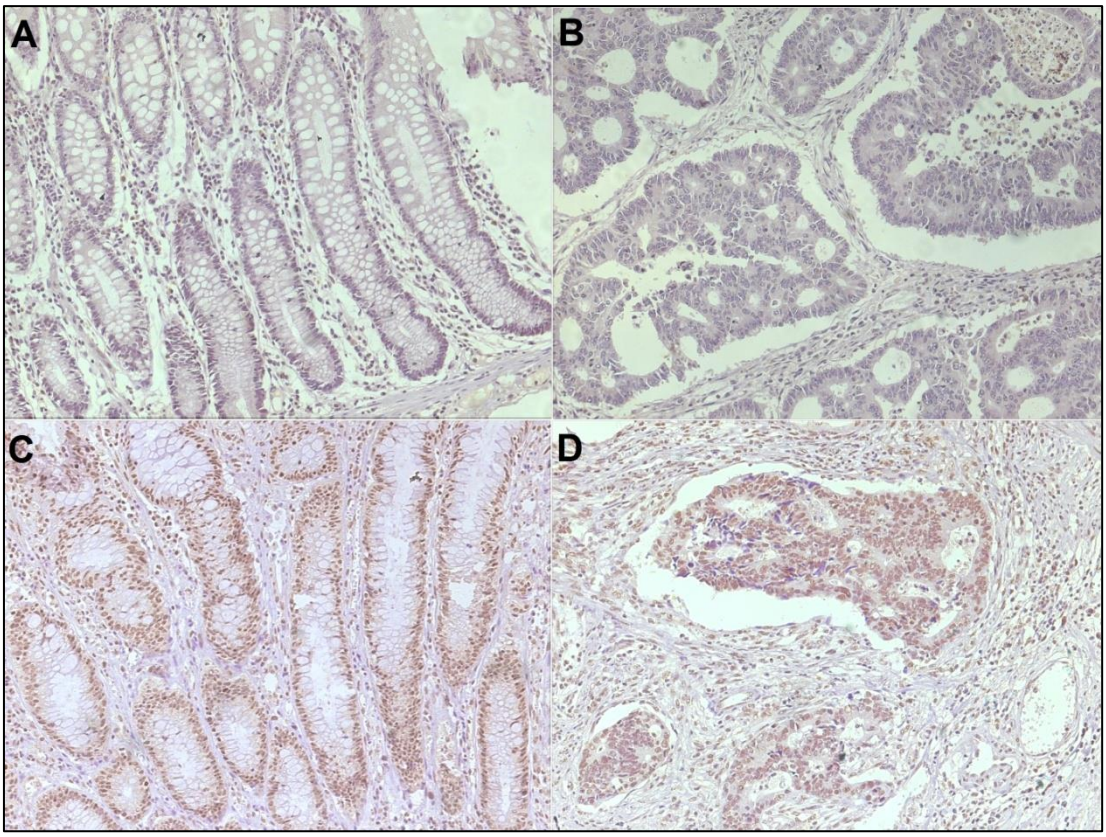


Figure 6.6. Examples of RPL6 immunostaining in samples stored for different lengths of time. The IHC protocol from the Sao Nicolau lab was used. Images A and B show the staining pattern in a CRC case stored for 12 years (adjacent and cancer tissues, respectively). No immuno-reaction is observed either in epithelial or stromal cells. Images C (adjacent) and D (tumour) show a clearly positive nuclear staining in epithelial cells in a sample stored for two years. Magnification: 200x.

These results suggest that the protocol used is adequate for detecting RPL6 immunohistochemical expression in samples that are less than two years old, but not in those that have been stored for longer periods of time, e.g. the patient samples from the prognostic cohort. The fact that NAP1L1 and PHB (next section) were easily identified in this collection of older samples suggests that only RPL6 immuno-expression is affected by this degree of sample storage.

The most obvious differential characteristic between these two sample cohorts is the length of storage. However, stable antigenicity in FFPE blocks has been documented up to 68 years after sample collection (Bass *et al*., 2014). A few proteins have however been proven to show unstable antigenicity during prolonged storage, such as PCNA (Malmstrom *et al*., 1992). Antigen degradation is much more common in slides that have been

cut months to years prior to IHC experiments (Grillo *et al*., 2015). This was not the case in our investigations since we only used slides that had been cut from the paraffin block days to weeks prior to the experiments. In the specific case of the experiments shown in figure 6.6 above, we used slides that had been prepared the previous day. RPL6 stability in stored slides or FFPE blocks has not been previously examined. The samples used in the staining comparison had been stored for only one year when the initial validation experiments took place (figure 6.5A), whilst they were 3 years-old when the prognostic study was performed (figure 6.5B). Therefore, an effect of the length of storage on RPL6 immune-reaction, although unlikely, cannot be excluded from our data.

Other ambient conditions (such as temperature and humidity) have also been linked to decreased immunogenicity in archival slides (Xie *et al*., 2011). Again, the effect of these parameters on paraffin blocks is less clear. Given that the blocks used in the prognostic study were retrieved from pathology services which did not have standardised protocols to control storage conditions, it is possible that these factors may have influenced the preservation of the samples. Of note, temperatures in Cuiaba, the Brazilian city were this collection took place, can reach daily averages above 30°C during the warmest months of the year. Additionally, standard operational procedures may have changed over time in terms of pre-analytical factors such as pre-fixation, fixation and processing conditions. Therefore, it was not possible to fully explain the precise cause of the negative results observed. Nonetheless, it is clear that the lack of reaction was restricted to RPL6. Issues related to antibody affinity especially in long term stored samples were also a possible cause for the negative staining, as the other antibodies tested showed adequate results. The use of different RPL6 antibodies could potentially clarify this point. Unfortunately such additional experiments could not be carried out due to time constraints.

As the samples that yielded adequate immune reactions all had limited follow up times (less than three years), a survival study could therefore not be performed. Additionally, clinical information for these cases was not available, as this was not the intended cohort for the prognostic study.

Therefore, we could not adequately assess the prognostic role of RPL6 immuno-expression in CRC patients. Further studies to assess RPL6 expression in samples that have been stored under controlled conditions could unveil its prognostic value or any possible effect of storage conditions on protein degradation.

6.5. PHB as a prognostic marker in CRC

The same stepwise approach described for NAP1L1 was adopted for the assessment of PHB as a prognostic marker in CRC. PHB immunostaining was similar using both staining methods (figure 6.7). Nuclear and cytoplasmic scores were analysed using ROC curves (AUCs 0.59 and 0.57, respectively). Scoring cut-offs were set at 18 (percent) for nuclear scoring (sensitivity: 61.5%; specificity: 57%) and at 110 for cytoplasmic scoring (sensitivity: 54%; specificity: 50%). After splitting the samples into low and high expression groups, the first analysis was the comparison between groups in relation to clinicopathologic variables in the same way as described for NAP1L1 (see table 6.2 for reference). As was the case for NAP1L1, the groups were well balanced. Neither the nuclear nor the cytoplasmic expression of PHB was associated with age, gender, grade or stage (table not shown).

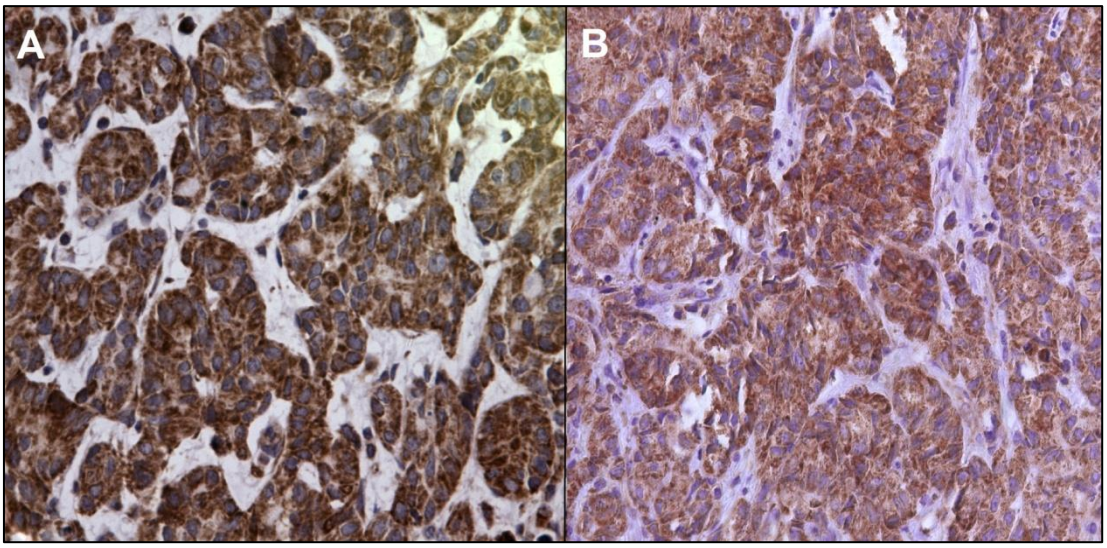


Figure 6.7. Comparison of PHB staining patterns observed with the IHC protocols used in the initial validation study (A) and in the prognostic study (B). Similar staining intensity and localisation were observed. Differences in staining shade and background were due to the use of different cameras for recording the images (magnification: 400x).

A survival analysis using the same method employed for NAP1L1 is shown in figure 6.8. Considering all stages combined, survival was very similar until around 2 years of follow-up. After this time point, an apparent separation of the survival curves was observed. However, this difference was not statistically significant (p=0.286, log-rank test). Estimated 5-year survival was 67% for the low expression group and 57% for the high expression group; this was also non-significant. For the cytoplasmic staining, survival curves were similar during the entire follow up period. The same was observed in terms of 5-year survival estimates: 60% and 65% for the low expression and the high expression cohorts, respectively (p=0.769, log-rank).

Analyses were also carried out for early and late stage disease separately and, similarly, no significant differences were observed, thus confirming the initial findings (figure 6.9). Therefore, no significant impact of PHB protein expression on prognosis was found. This finding is partially concordant with the results from Cheng *et al*. who reported no association between PHB expression and survival in CRC patients, although a correlation with tumour grade was observed in that study (Chen *et al*., 2010a). We are unaware of any other report that has assessed the role of PHB immunostaining in affecting the prognosis of CRC patients.

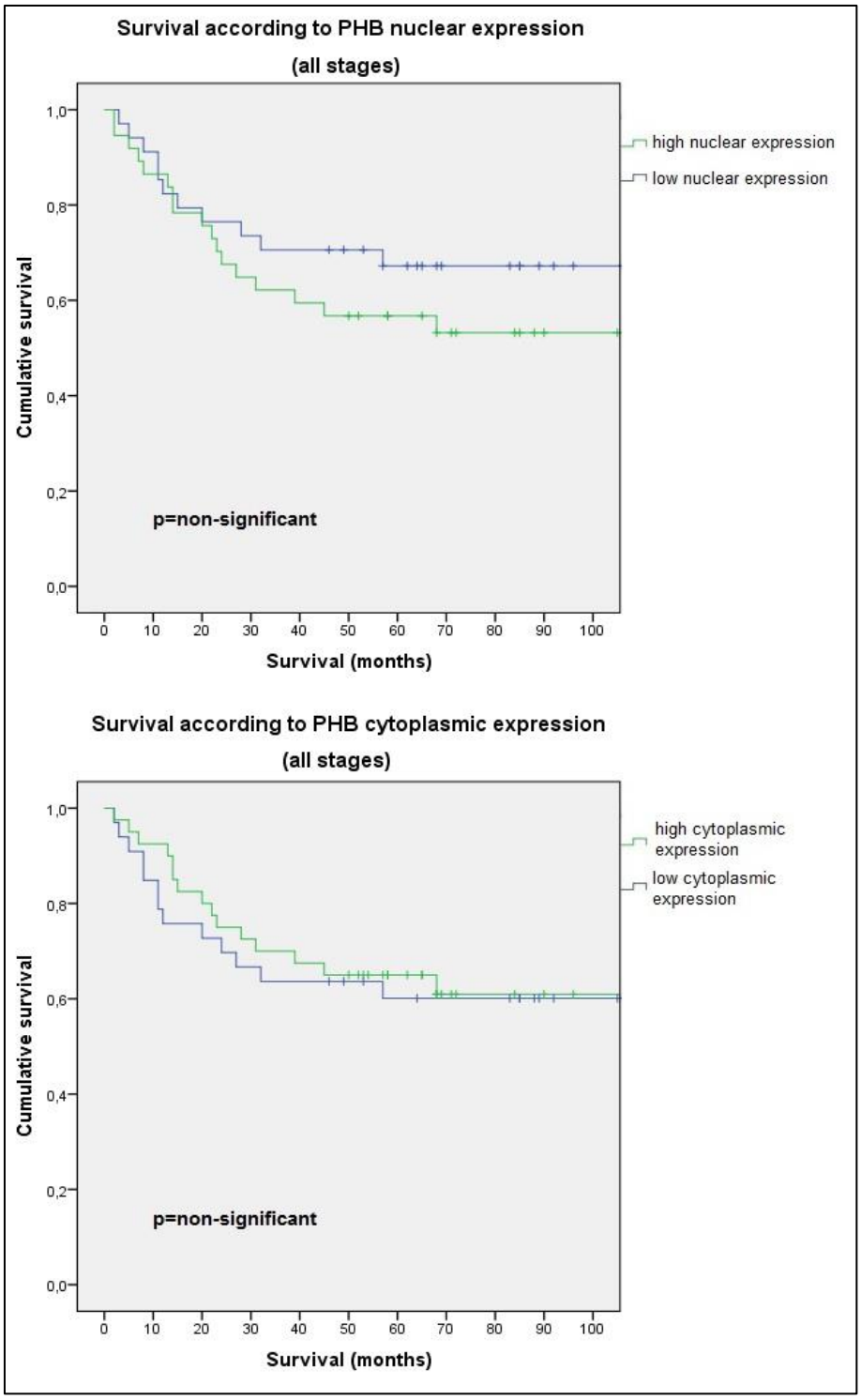


Figure 6.8. Cumulative survival according to PHB expression (all stages combined). The analyses of both nuclear and cytoplasmic immunostaining patterns did not reveal significant differences in survival when comparing low and high expression groups. However, a trend towards improved survival was observed in the low nuclear expression group. Vertical marks=censored cases.

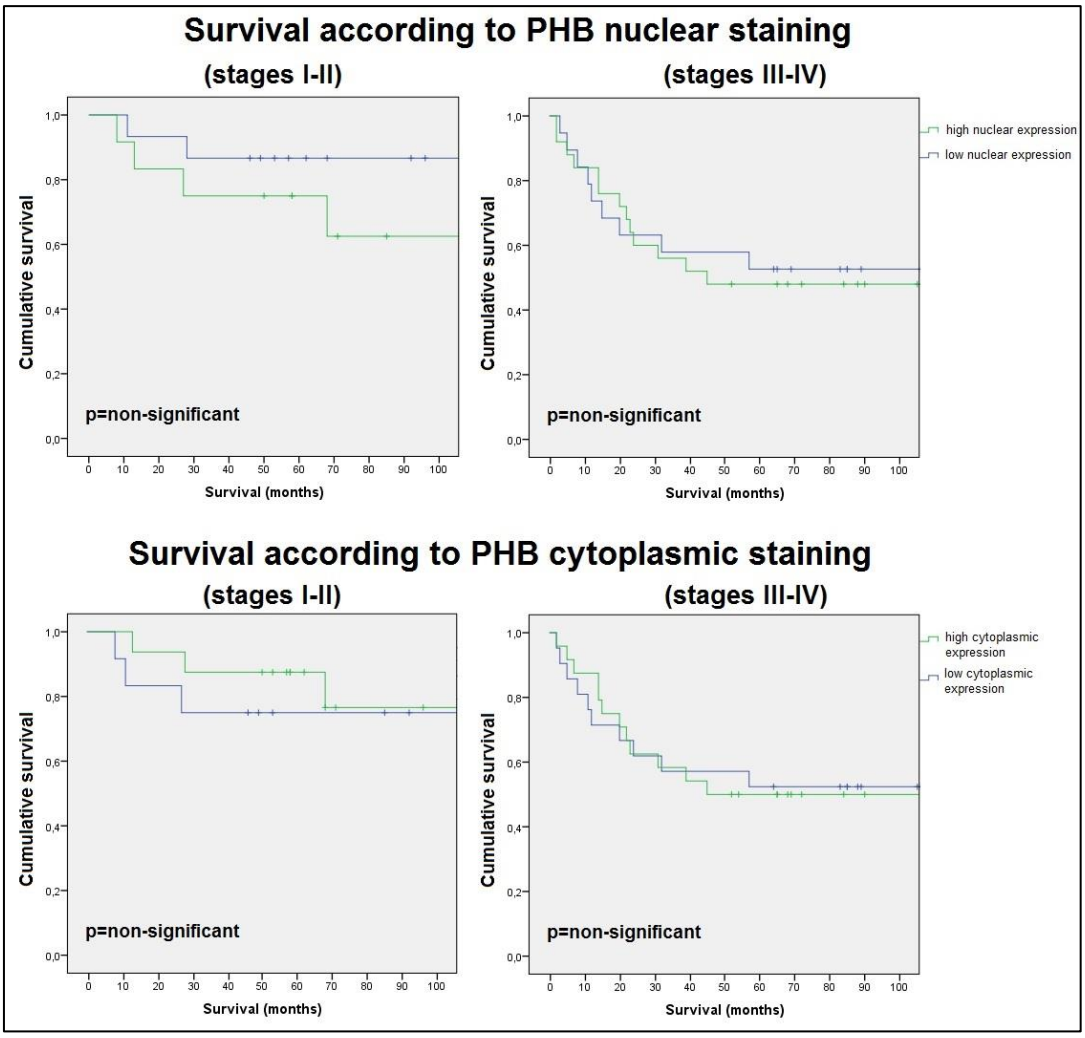


Figure 6.9. Cumulative survival according to PHB staining in different stage groups. No significant difference between low and high expression groups was observed regardless of the tumour stage. Of note, a trend of improved survival was observed in the low nuclear expression group, for early stage tumours only.

6.6. Discussion

Prognostic markers are helpful for treatment decision-making. Individuals with a higher risk of recurrence or death caused by CRC may benefit from more aggressive therapeutic approaches or from more intensive surveillance programmes after treatment. Conversely, low-risk patients could potentially be spared from more radical surgery or toxic adjuvant therapies. Cancer stage is the most extensively studied and validated prognostic factor in CRC. Several other histopathological and molecular markers also seem to have prognostic significance, although none is as consistent as tumour stage (Sagaert, 2014). Nonetheless, it has long been recognised that TNM

(tumour-node-metastasis) and other similar staging systems (such as the Dukes') are not perfect prognostic indicators, since diseases within similar stage groups may exhibit different long-term outcomes (Burke, 2004). Advanced stage may be the consequence of worse tumour biology rather than the cause of higher mortality *per se*. This discussion is particularly relevant for stage II tumours, which are deemed low-risk disease and, as such, patients are usually not offered post-surgical adjuvant chemotherapy (Benson *et al*., 2004). However, 5-year survival among stage II patients is not optimal – usually between 70-80% (Morris *et al*., 2006, Petersen *et al*., 2002), with some studies reporting survival as low as approximately 50% depending on tumour depth, number of lymph nodes resected, presence of vascular invasion or peritoneal involvement and postoperative CEA levels (Peeples *et al*., 2010, Quirke and Morris, 2007, Petersen *et al*., 2002). A systematic review has suggested that some stage II patients might actually benefit from additional treatments (Wu *et al*., 2012b). Although this controversy (whether or not to offer adjuvant chemotherapy) is valid for stage II tumours, all stage groups exhibit diverse long-term outcomes and might also benefit from better prognostic determination.

Based on the idea that biology might be more relevant than morphology (Burke, 2004), molecular prognostic biomarkers derived from pathways associated with colorectal carcinogenesis have been extensively pursued. Mutations or methylation of several oncogenes and tumour suppressors are associated with prognosis in CRC patients (Yiu and Yiu, 2016, Erstad *et al*., 2015). Several non-coding RNAs have also been linked to disease progression and survival (Yiu and Yiu, 2016, Saus *et al*., 2016). Proteomics has also been used for biomarker discovery in CRC (Alvarez-Chaver *et al*., 2014). Although a plethora of candidate biomarkers has been suggested, no protein has been adopted as a prognostic biomarker mostly due to inconsistencies in reproducibility and lack of proper validation studies (Alvarez-Chaver *et al*., 2014, Coghlin and Murray, 2015).

The immunohistochemical expression of β-catenin has been widely studied, but with inconsistent results. A meta-analysis involving 3665 cases from 18 studies has demonstrated that the nuclear expression of β-catenin is

associated with a worse prognosis in CRC patients (Chen *et al*., 2013). Therefore, we decided not to assess the prognostic value of this protein in our research. Instead, we concentrated our efforts on the analysis of the novel potential biomarkers that had been suggested in our previous studies. Using a cohort of 75 Brazilian CRC cases who had a relatively long follow up, we assessed NAP1L1, RPL6 and PHB expression and correlated these markers with clinicopathologic variables. Furthermore, survival correlations were performed in order to define their prognostic significance.

The importance of NAP1L1 immunohistochemical expression in relation to prognosis has not been previously reported either in CRC or in any other cancer type. Therefore, our study provides the first evidence for a role of this biomarker in predicting clinical outcome in CRC patients. The experiments described in this chapter clearly suggest that the nuclear expression of NAP1L1 is related to overall prognosis. High nuclear expression was independently associated with a marked increase in survival duration and 5-year survival estimates. Mortality in this group was 61 - 72% lower when compared with the low-expression group. Subgroup analyses showed that the survival correlation was however limited to late stage tumours (stages III and IV). No association between NAP1L1 nuclear expression and clinicopathologic variables (age, gender, stage and grade) was observed. As a result, we conclude that NAP1L1 nuclear immunostaining is a strong survival predictor in CRC patients. However, NAP1L1 cytoplasmic expression did not correlate with overall survival or any other clinicopathologic feature.

RPL6 prognostic importance has not previously been studied in human cancers. Our attempt to assess the immuno-expression of this protein as a prognostic factor in CRC was unsuccessful mainly due to inadequate immune reaction in samples which had been stored for long periods of time. The IHC procedures performed were deemed suitable for this assessment, as cases which had shorter follow up times produced positive immunostaining. The reason for the absence of immune reaction using the older samples was not clear and requires further investigation.

The prognostic significance of PHB immuno-expression has previously been studied in various cancer types. The expression of this protein is reduced in poorly differentiated ovarian cancers when compared with more differentiated tumours (Jia *et al*., 2014). A decrease in PHB immunostaining has also been correlated with worse prognosis in nasopharyngeal tumours (Liao *et al*., 2013). Conversely, increased PHB immunostaining has been associated with adverse prognostic features in lung (Jiang *et al*., 2013, Guo *et al*., 2012), breast (Najm *et al*., 2013), gastric (Kang *et al*., 2008), thyroid (Franzoni *et al*., 2009), prostate (Ummanni *et al*., 2008) and bladder cancers (Wu *et al*., 2007, Cao *et al*., 2016). The only study that we found assessing the role of PHB immuno-expression in CRC reported no correlation with overall survival (Chen *et al*., 2010a). In the present study, we did not observe a significant association between PHB expression and overall survival in CRC patients. Neither nuclear nor cytoplasmic PHB immunostaining was associated with the clinicopathologic characteristics aforementioned. Similarly, no significant survival association was found, a finding concordant with the results of Chen *et al*., 2010a. Only a non-significant trend towards improved survival in the low nuclear expression group was noticed and this trend was limited to early stage cancers. Therefore, our data do not support the hypothesis that PHB immunohistochemical expression is a prognostic biomarker in CRC.

We recognise that this study has several limitations. The limited sample size did not allow us to fully assess the prognostic importance of the biomarkers for early stage cancers, especially stage II tumours. This was caused by the low number of events (deaths) in these groups. Therefore, it would be highly desirable to replicate this study in a larger cohort of early stage cancers in order to definitively assess whether or not this biomarker could be used for the identification of high-risk early stage disease. Additionally, for a number of cases, we could not determine precisely whether the deaths that occurred were caused by CRC (*CRC-specific mortality*) due to the fact that the *Mortality Information System* lists only the direct cause of death, regardless of the underlying condition. For example, a patient who died from pneumonia as a result of cancer progression would

have been registered as “dying from pneumonia” even when this outcome was clearly related to cancer progression. However, overall survival is the “gold standard” clinical endpoint in cancer studies (Sherrill *et al*., 2012) and this reinforces the importance of our findings. Another outcome we would like to have explored in this study was CRC recurrence. Unfortunately, this was not possible due to the lack of information about this for most patients, but this is certainly a point to consider if further studies are undertaken. Tumour recurrence is an earlier event requiring a shorter follow up period to be properly assessed and is thought to be an adequate surrogate endpoint for overall survival in CRC (Giessen *et al*., 2013). If available, this information could also unveil a prognostic impact in stage I and II tumours. Additionally, we were not able to obtain accurate information regarding the adjuvant treatments that patients had received. Hence, the presence of a treatment misbalance between groups could not be assessed. In terms of the technical procedures performed, although objective scoring methods were used, manual and inter-observer agreement was not assessed. Most pathology labs use manual rather than electronic scoring and agreement between professionals, especially less experienced ones, may be an issue (Ali *et al*., 2016, Jaraj *et al*., 2009). Despite these caveats, the magnitude of the impact of NAP1L1 nuclear immunostaining on the prognosis of CRC patients strongly suggests that there is a real correlation. This might therefore be a valuable tool to define patients who have a poorer prognosis and may in future help guide treatment or follow up decisions. Comprehensive studies involving larger populations and other prognostic endpoints are however necessary to definitely confirm the clinical utility of this novel CRC biomarker before its use can be routinely recommended.

Chapter Seven:
General discussion

7. CHAPTER 7 – GENERAL DISCUSSION

The importance of finding CRC biomarkers for early diagnosis, prognostic determination and prediction of response to treatment cannot be underestimated. CRC is the third leading cause of cancer mortality in men and the second in women worldwide (Ferlay *et al*., 2015) and most deaths result from late diagnosis. A better understanding of disease biology and the discovery of accurate biomarkers may permit the use of therapies that are tailored for individual patients – personalised medicine. Despite the multitude of potential candidates, very few CRC biomarkers have yet been adopted into routine clinical practice (Coghlin and Murray, 2015). Therefore, further intensive translational research is still necessary in order to transfer innovative tools from the laboratory bench to the hospital bedside.

The use of prospection strategies based on solid biological rationale is more likely to yield useful biomarkers than non-targeted approaches such as the simple comparison of the whole “omics” profiles of normal versus malignant tissues. Effective forms of biomarker discovery rely on the exploration of molecular pathways that are associated with carcinogenesis. The Wnt signalling pathway is up-regulated in the majority of sporadic CRC cases. *APC* mutation (leading to Wnt activation) is the most common genetic event in the early steps of the normal-adenoma-carcinoma transformation sequence (Cancer Genome Atlas Network, 2012). Previously in our group, Hammoudi (Hammoudi *et al*., 2013) and Ibrahim (Ibrahim, 2014) demonstrated that the analysis of animal models of acute and chronic *Apc* inactivation led to the selection of potential candidate biomarkers of human colorectal carcinogenesis. In the research described in this thesis, we have performed a comprehensive analysis of these candidate biomarkers using scientific techniques that are widely available in clinical and research laboratories. Whenever possible, we have assessed the expression of these potential biomarkers in different patient populations in order to investigate the potential effect of regional differences (ethnic variation, collection and storage conditions) on the results. Using this strategy, we have proven the validity of some candidate biomarkers and have confirmed that the use of animal

models of CRC based on *Apc* inactivation is a useful approach for biomarker prospection. Individual results have been comprehensively discussed in each dedicated chapter. A general discussion of the main findings is provided below.

7.1. Confirmation of differential immuno-expression of the biomarkers

Immunohistochemistry is an inexpensive, widely available technique and is routinely used to assess cancer biomarkers in several tumour types (Chamberlain *et al*., 2015, Zaha, 2014, Toffart *et al*., 2014, Varma and Jasani, 2005). In CRC patients, IHC has been used for assessing MMR proteins for the diagnosis of HNPCC and for prognostic determination in sporadic tumours (Steinhagen *et al*., 2012, Yoon *et al*., 2011). No immunohistochemical marker has however yet been routinely adopted for clinical use in CRC. Using electronic scoring methods, Wnt activation was initially confirmed in all neoplastic tissues, as nuclear and cytoplasmic over-expression of β-catenin was observed in low-grade adenomas, high-grade adenomas and cancer tissues. These results concur with the published literature (Wong *et al*., 2004, Wong *et al*., 2003, Chen *et al*., 2013). We then proceeded with the evaluation of our novel candidate biomarkers. A clear differential expression was observed for NAP1L1, RPL6 and PHB when comparing tumours and non-malignant tissues. This confirmed the findings from the animal work that has previously been conducted in our research group. HMGB1, SFRS2 and CDC5L did not exhibit different immunostaining patterns. As a result, these proteins were not further assessed in this thesis.

NAP1L1 and RPL6 demonstrated similar staining patterns in all sample groups analysed. Both proteins exhibited a clear decrease in nuclear expression in neoplastic tissues compared to the adjacent normal colonic mucosa. A similar, although less pronounced, difference was also observed for NAP1L1 cytoplasmic expression. Additionally, the expression of NAP1L1 and RPL6 in low-grade adenomas was comparable to that observed in adjacent normal mucosa and normal control samples, whilst adenomas with

high-grade dysplasia exhibited a staining pattern more similar to cancer tissues. This latter finding provides a rationale for studying NAP1L1 and RPL6 expression to better discriminate between high-risk and low-risk adenomas, a possibility that still requires further research. To the best of our knowledge, no study of the immuno-expression of these candidate biomarkers in CRC has previously been published.

PHB nuclear expression was also decreased in malignant colorectal tissues. However, the cytoplasmic expression of this protein exhibited a slight increase in cancer samples. This finding suggests that the protein might have been displaced from the nucleus towards the cytoplasm during malignant transformation. An increase in PHB cytoplasmic content has already been reported in CRC (Chen *et al*., 2010a), but this study did not assess its nuclear expression. Taken together, these findings confirm the potential of NAP1L1, RPL6 and PHB immuno-expression as biomarkers for CRC development. Validation studies involving larger populations may unveil the roles of these proteins for cancer screening and diagnosis.

7.2. Gene expression studies suggest a role for *NAP1L1* and *RPL6* in field cancerisation

Messenger RNA expression analysis using qPCR has been widely used to study the expression of genes involved in colorectal carcinogenesis, although methodological rigour has frequently been poor (Dijkstra *et al*., 2014). Despite this, commercial tests using gene expression panels have been developed such as *Oncotype Dx Colon*™ (Clark-Langone *et al*., 2010) and *ColoPrint*™ (Kopetz *et al*., 2015). Both tests aim to assess the risk of recurrence in localised tumours. Prognostic tests such as these have not gained much acceptance by clinicians, mainly due to the lack of proper validation in large population settings (Webber *et al*., 2010). Robust transcriptomic biomarkers are therefore still lacking in this disease.

We performed a qPCR-based analysis of the expression of *NAP1L1, RPL6* and *PHB* in order to further validate them as potential biomarkers in CRC. Additionally, we tested *CTNNB1* alongside to assess the function of this gene in this biological scenario. We analysed cancer samples, adjacent unaffected mucosa and normal control mucosa (from individuals who did not have colorectal cancer or polyps).

The expression of *CTNNB1* was only slightly increased in cancer tissues compared to both the adjacent and the normal control mucosa. The literature reports on the expression of this gene in CRC are conflicting (Anwar *et al*., 2015, Qin *et al*., 2006, Truant *et al*., 2008). Our finding was therefore not unexpected. Activation of the Wnt pathway in CRC is more commonly a result of *Apc* mutations resulting in accumulation of β-catenin via decreased degradation than as a consequence of increased *CTNNB1* mRNA expression. Even when *CTNNB1* is the genetic driver of carcinogenesis, mutations, rather than increased expression, appear to be the relevant alteration (Cancer Genome Atlas Network, 2012).

Regarding our candidate biomarkers, we again observed a strikingly similar pattern of expression for *NAP1L1* and *RPL6*. Both genes were shown to be highly expressed in the tumour and the adjacent mucosa from patients with CRC when compared to normal controls. “Field cancerisation” or “field effect”, may explain these findings. It has been well described in cancer development mostly as a result of genetic and epigenetic modifications leading to altered gene expression (Baba *et al*., 2016), making tissues adjacent to tumours prone to developing new primary cancers. It has been shown to occur up to several centimetres from the primary colorectal tumour (Park *et al*., 2016). Therefore, even taking into account that the “adjacent” samples were collected 10cm from the primary tumour in our study, this event could still have taken place.

On the other hand, differences in patient populations or sample collection procedures might have affected these results. There were more men in the cancer group (who provided both tumours and adjacent tissues) than in the normal control group. However, no influence of gender or sex

hormones on the expression of *NAP1L1* or *RPL6* that could explain the difference in gene expression has previously been reported. Additionally, factors related to the procedures performed for sample collection might explain the findings. Bowel preparation and type of anaesthesia are different for colonoscopy and for surgical colectomy and this difference may have had an effect on the general gene expression profile. Nonetheless, the fact that the expression of *CTNNB1* was similar between adjacent and normal controls weakens this explanation as the cause of the observed difference.

PHB gene expression results were not statistically significant, although a trend pointing towards increased transcript levels in adjacent normal and tumour tissues was observed. The results from Hammoudi *et al*. showed only a 1.32-fold increase in *PHB* expression in tumour versus adjacent samples (Hammoudi *et al*., 2013). Along with our findings, this suggests that *PHB* mRNA over-expression might not be as important in CRC as it has been demonstrated in other cancer types (Franzoni *et al*., 2009, Jiang *et al*., 2013, Kang *et al*., 2008, Ummanni *et al*., 2008). We could not find any other study that has assessed *PHB* gene expression in CRC.

7.3. *RPL6* siRNA silencing results in inhibition of proliferation and altered expression of cancer-associated genes in CRC cells

After observing a differential expression of *NAP1L1* and *RPL6* in cancer and adjacent tissues, we decided to explore the mechanisms by which these genes affect cancer cell biology. For this purpose, we carried out siRNA experiments aiming to knockdown the expression of these genes in *TP53* wild-type and null HCT116 colorectal cancer cells. After successful transfection of these cells, a proliferation assay (SRB) was performed. Cells without a functioning *TP53* gene exhibited severe toxicity probably caused by the transfection reagents and/or by nutrient deprivation during the proliferation assay. Therefore, the results observed with these cells were not deemed sufficiently accurate to draw any meaningful conclusions.

In *TP53* competent cells, *NAP1L1* silencing did not cause any alteration in cell proliferation. The literature suggests that *NAP1L1* regulates the differentiation and proliferation of stem cells (Li *et al*., 2012, Gong *et al*., 2014, Yan *et al*., 2016) and pancreatic neuroendocrine tumours (Schimmack *et al*., 2014). However, no such suggestion has been made regarding other cancer cells. Therefore, disturbed *NAP1L1* expression in CRC might not be associated with proliferation but, rather, be essential for the maintenance of an undifferentiated status.

Differently, *RPL6* silencing resulted in strong inhibition of cell growth in HCT116 cells. Proliferation inhibition upon *RPL6* silencing has already been demonstrated in gastric cancer cell lines (Wu *et al*., 2011, Du *et al*., 2005). Here, we provide the first evidence that this gene also affects the proliferation of a CRC cell line. As the mechanistic functions of *RPL6* in the regulation of cell proliferation have been poorly described, we decided to assess whether other genes related to colorectal carcinogenesis would be affected by *RPL6* knockdown. For this, we analysed the expression of a panel of 28 CRC-associated genes using RNA extracted from cells transfected with *RPL6* siRNA or scrambled siRNA. Using this strategy, we demonstrated that *RPL6* silencing resulted in the up-regulation of *BAX* and *MSH2* (apoptosis and mismatch repair regulators, respectively) and down-regulation of *MMP-12* and *MMP-13* (metalloproteinases involved in invasion, angiogenesis and metastasis). Collectively, these results strongly support the hypothesis that *RPL6* is associated with the development of malignant hallmarks in CRC. Further research is necessary to confirm and clarify these findings.

7.4. Biomarker concentrations are not consistently increased in blood from individuals with CRC or adenoma

Blood tests are favoured as minimally-invasive and widely accepted forms of obtaining biological samples (Rifai *et al*., 2006). In CRC screening, a study has suggested that a blood test would result in higher acceptance by the population than colonoscopy (Adler *et al*., 2014). Protein biomarker

discovery usually involves the initial use of non-targeted approaches (such as LC-MS/MS) exploring the whole proteomic profile of samples, followed by a more targeted confirmatory phase usually based on enzyme immunoassays such as ELISA (Rifai *et al*., 2006).

In this part of the research, immunoassays assessing NAP1L1, RPL6 and PHB as potential CRC blood biomarkers were performed. Sample cohorts from Brazil and from the UK were used in order to generate robust and valid results. Plasma and serum were tested, as neither of these fluids had been suggested as the optimal matrix for the measurement of those proteins. No ELISA kit had previously been confirmed as the gold standard for any of the candidate proteins. Therefore, we initially tested kits from different manufacturers in order to assess whether they produced uniform and concordant results.

NAP1L1 proved to be the most problematic protein in the immunoassays. All three NAP1L1 ELISA kits tested (from Cloud Clone Corp, DL Develop and Abbexa) exhibited poor performances and produced unreliable results. The only consistent finding was the observation that plasma NAP1L1 concentrations were usually higher than serum concentrations. As a result of these unsatisfactory findings, we developed an in-house immunoassay based on the Meso Scale Discovery® electrochemiluminescence (ECL) platform, a method which has results generally superior to ELISAs (Chaturvedi *et al*., 2015, Sloan *et al*., 2012, Postelnek *et al*., 2016). Quality control tests confirmed that this novel ECL assay was more consistent and sensitive than the ELISAs previously used. This assay confirmed that plasma was the more adequate matrix for NAP1L1 assessment. Analyses of the Brazilian and UK cohorts using this in-house immunoassay however did not demonstrate significant differences in NAP1L1 concentrations between normal controls, adenoma-bearing individuals and CRC patients. A trend towards increased plasma NAP1L1 concentrations in high-grade dysplastic adenomas and cancer samples was observed, but statistical significance was not reached.

RPL6 and PHB ELISA kits demonstrated better performances than NAP1L1 kits in terms of accuracy and consistency. Plasma was again the chosen matrix due to the higher protein concentrations observed. Similarly, for both candidate markers, the assessment of protein concentrations did not show any significant differences between clinical groups. These results are different from the initial findings observed by our research group. The very limited sample size used in the preliminary phase may however have been the cause of this difference. Our final tests were performed using larger and more diverse sample cohorts. The consistent results exhibited by plasma samples allowed us to conclude that both RPL6 and PHB are not differentially expressed in the blood of individuals with CRC or adenomas when compared with normal individuals.

Although we fulfilled the objective of developing an adequate immunoassay for NAP1L1, this work did not succeed in identifying any of these proteins as potential CRC blood biomarkers. However, important new insights related to the study of these proteins in blood-derived fluids were generated. Furthermore, a possible effect of sample collection procedures or storage conditions on protein levels was unveiled, as we found differences in results between similar groups from Brazil and the UK. These findings may have important implications for any future research assessing the concentrations of NAP1L1, RPL6 or PHB in biological fluids.

7.5. NAP1L1 nuclear expression is a strong predictor of survival in late stage CRC

Prognostic biomarkers are defined as clinical or laboratory characteristics that are capable of predicting clinical outcome such as tumour recurrence and, most importantly, survival (Pritzker, 2015, Atkinson *et al*., 2001, Sherrill *et al*., 2012). Despite the comprehensive search for CRC prognostic biomarkers and the several candidates suggested in the literature, very few of these have been incorporated into routine clinical practice

(Bianchi *et al*., 2011, Watson and Søreide, 2016, Sagaert, 2014). Here, we performed an assessment of the relationship between the immunohistochemical expression of our candidate proteins, prognostic variables and patient survival.

The experiments carried out clearly demonstrated that the nuclear expression of NAP1L1 was related to prognosis in patients with CRC. High nuclear expression was independently associated with a marked increase in survival duration and 5-year survival estimates. Mortality in this group was 61 to 72% lower when compared with the low-expression group. Subgroup analyses suggested that the survival correlation was limited to late stage tumours (stages III and IV). However, the limited number of events (deaths) may have prevented us from finding a correlation for early stage, particularly stage II, tumours. No association between NAP1L1 nuclear expression and age, gender, grade or stage was observed. The prognostic importance of NAP1L1 has not been previously reported in CRC or other cancer types. Thus, the present study provides the first evidence for an association between the expression of this protein and clinical outcomes in cancer patients. Noteworthy, the cytoplasmic expression of NAP1L1 was not associated with any prognostic factor in this study.

We did not observe a significant relationship between PHB expression and prognostic factors in CRC patients. Neither nuclear nor cytoplasmic PHB immunostaining was associated with clinicopathologic characteristics or survival. This finding is concordant with the results from Chen *et al*. who also assessed PHB immuno-expression in CRC patients. Consequently, it is our conclusion that PHB immunohistochemical expression is not a prognostic biomarker in CRC.

RPL6 assessment was impaired by the absence of immune reaction in the IHC experiments performed using the prognostic cohort of samples. This negative reaction seemed to be specifically associated with this group of samples, since cases with shorter storage times exhibited adequate immunostaining. It is not clear whether this difference was caused by the length of storage *per se* or by any other pre-analytical factor.

7.6. Study limitations and suggestions for future research

The use of human clinical samples for research is intrinsically problematic. Different from animal models or cell lines, it is difficult to obtain tissue or blood from cancer patients in large quantities unless a high-volume cancer hospital is committed to the work or a multi-centric effort is assembled. Even in those situations, the prospective collection of samples still requires a reasonably long time. In this work, samples were collected in small cancer centres in Brazil and in a single district general hospital in the UK. Additionally, problems with sample transportation from Brazil to the UK caused the loss of large amounts of biological materials. As a result of these issues, sample sizes used in several parts of this research were suboptimal. Although several positive and relevant findings were made, the use of small sample sizes may have impeded the observation of more positive results in particular sample cohorts. Another consequence of the limited number of samples was the necessity of analysing cancer stages as groups and not individually. It would in future be interesting to assess the stages separately. Better prognostic stratification in stage II disease is desperately needed to improve the treatment decision-making process. Besides, splitting stage III and stage IV diseases would also be desirable, as these stages result in markedly different clinical outcomes.

Regarding the IHC scoring method, objective electronic systems were used in this study. These methods have the advantage of improving result consistency especially for non-pathologists. However, most pathology services do not use electronic scoring tools, but manual scoring methods instead. Therefore, it would be important to assess the reproducibility of our findings via manual scoring performed by trained pathologists, ideally by more than one professional.

In the immunoassay experiments, an unanswered question is the influence of sample storage conditions upon protein levels. Thus, the assessment of sample stability by testing different lengths and temperatures of storage, and several freeze-thaw cycles would provide guidance for future studies assessing the concentration of the proteins in biological fluids.

In the prognostic study, apart from the sample size issue already mentioned, another drawback was the lack of data about treatments that had been performed and tumour recurrence. This information would provide more insights into the significance of the biomarkers in the management of CRC patients.

7.7. Conclusions

Despite the issues discussed above, this study has produced several interesting and consistent findings. It confirmed that NAP1L1, RPL6 and PHB are proteins that are differentially expressed in human colorectal neoplasms, as suggested by the initial animal work performed by previous members of our research group. It also demonstrated that the gene expression of *NAP1L1* and *RPL6* is increased in tumours and in the adjacent mucosa from CRC patients, suggesting that these biomarkers may be involved in tumourigenesis and field cancerisation. Mechanistic studies showed an effect of *RPL6* expression on the proliferation of CRC cells. Additionally, the silencing of this gene caused an alteration in the expression of other cancer-associated genes, reinforcing its importance in colorectal carcinogenesis. The assessment of the prognostic impact of the expression of the candidates demonstrated a strong effect of NAP1L1 nuclear immunostaining on the survival of patients with late stage disease, a finding which has potential clinical applications. The analysis of protein levels in blood samples, however, did not show increased concentrations of any of the postulated biomarkers in patients with either adenomas or colorectal cancer.

In summary, this research has confirmed the validity of the study of animal models of *Apc* inactivation for the prospection of CRC biomarkers. It has suggested that NAP1L1, RPL6 and PHB are involved in colorectal carcinogenesis and may be useful biomarkers for the diagnosis of high-risk adenomas and CRCs. In addition, it has unveiled a possible role of NAP1L1 as a prognostic marker in this disease. To avoid these candidates being placed in the long list of promising, but non-validated biomarkers of CRC,

further prospective research assessing larger sample cohorts is highly recommended.

8. REFERENCES

ABDELMAKSOUD-DAMAK, R., MILADI-ABDENNADHER, I., TRIKI, M., KHABIR, A., CHARFI, S., AYADI, L., FRIKHA, M., SELLAMI-BOUDAWARA, T. & MOKDAD-GARGOURI, R. 2015. Expression and mutation pattern of beta-catenin and adenomatous polyposis coli in colorectal cancer patients. *Arch Med Res,* 46**,** 54-62.

ADAMS, J. M. & CORY, S. 2007. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene,* 26**,** 1324-37.

ADLER, A., GEIGER, S., KEIL, A., BIAS, H., SCHATZ, P., DEVOS, T., DHEIN, J., ZIMMERMANN, M., TAUBER, R. & WIEDENMANN, B. 2014. Improving compliance to colorectal cancer screening using blood and stool based tests in patients refusing screening colonoscopy in Germany. *BMC Gastroenterol,* 14**,** 183.

AHLQUIST, D. A., SKOLETSKY, J. E., BOYNTON, K. A., HARRINGTON, J. J., MAHONEY, D. W., PIERCEALL, W. E., THIBODEAU, S. N. & SHUBER, A. P. 2000. Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay panel. *Gastroenterology,* 119**,** 1219-27.

AHNEN, D. J. 2011. The American College of Gastroenterology Emily Couric Lecture--the adenoma-carcinoma sequence revisited: has the era of genetic tailoring finally arrived? *Am J Gastroenterol,* 106**,** 190-8.

AJUH, P., KUSTER, B., PANOV, K., ZOMERDIJK, J. C., MANN, M. & LAMOND, A. I. 2000. Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J,* 19**,** 6569-81.

AL-DIMASSI, S., ABOU-ANTOUN, T. & EL-SIBAI, M. 2014. Cancer cell resistance mechanisms: a mini review. *Clin Transl Oncol,* 16**,** 511-6.

ALI, A., BELL, S., BILSLAND, A., SLAVIN, J., LYNCH, V., ELGOWEINI, M., DERAKHSHAN, M. H., JAMIESON, N. B., CHANG, D., BROWN, V., DENLEY, S., ORANGE, C., MCKAY, C., CARTER, R., OIEN, K. A. & DUTHIE, F. R. 2016. Investigating Various Thresholds as Immunohistochemistry Cutoffs for Observer Agreement. *Appl Immunohistochem Mol Morphol*.00:000–000 (Published Ahead of Print, POST AUTHOR CORRECTIONS).

ALI, R., BARNES, I., CAIRNS, B. J., FINLAYSON, A. E., BHALA, N., MALLATH, M. & BERAL, V. 2013. Incidence of gastrointestinal cancers by ethnic group in England, 2001-2007. *Gut,* 62**,** 1692-703.

ALLISON, J. E., FRASER, C. G., HALLORAN, S. P. & YOUNG, G. P. 2014. Population screening for colorectal cancer means getting FIT: the past, present, and future of colorectal cancer screening using the fecal immunochemical test for hemoglobin (FIT). *Gut Liver,* 8**,** 117-30.

ALSAIF, M., GUEST, P. C., SCHWARZ, E., REIF, A., KITTEL-SCHNEIDER, S., SPAIN, M., RAHMOUNE, H. & BAHN, S. 2012. Analysis of serum and plasma identifies differences in molecular coverage, measurement variability, and candidate biomarker selection. *Proteomics Clin Appl,* 6**,** 297-303.

ALTENBURG, F. L., BIONDO-SIMÕES, M. L. P. & SANTIAGO, A. 2007. Pesquisa de sangue oculto nas fezes e correlação com alterações nas colonoscopias. *Rev bras Coloproct,* 27**,** 304-309.

ALVAREZ-CHAVER, P., OTERO-ESTEVEZ, O., PAEZ DE LA CADENA, M., RODRIGUEZ-BERROCAL, F. J. & MARTINEZ-ZORZANO, V. S. 2014. Proteomics for discovery of candidate colorectal cancer biomarkers. *World J Gastroenterol,* 20**,** 3804-24.

AMUNDSON, S. A., MYERS, T. G., SCUDIERO, D., KITADA, S., REED, J. C. & FORNACE, A. J. 2000. An informatics approach identifying markers of chemosensitivity in human cancer cell lines. *Cancer Res,* 60**,** 6101-10.

ANASTAS, J. N. & MOON, R. T. 2013. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer,* 13**,** 11-26.

ANDERSON, N. L. & ANDERSON, N. G. 2002. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics,* 1**,** 845-67.

ANDREASSON, U., PERRET-LIAUDET, A., VAN WAALWIJK VAN DOORN, L. J., BLENNOW, K., CHIASSERINI, D., ENGELBORGHS, S., FLADBY, T., GENC, S., KRUSE, N., KUIPERIJ, H. B., KULIC, L., LEWCZUK, P., MOLLENHAUER, B., MROCZKO, B., PARNETTI, L., VANMECHELEN, E., VERBEEK, M. M., WINBLAD, B., ZETTERBERG, H., KOEL-SIMMELINK, M. & TEUNISSEN, C. E. 2015. A Practical Guide to Immunoassay Method Validation. *Front Neurol,* 6**,** 179.

ANKO, M. L. 2014. Regulation of gene expression programmes by serine-arginine rich splicing factors. *Semin Cell Dev Biol,* 32**,** 11-21.

ANWAR, M., KOCHHAR, R., SINGH, R., BHATIA, A., VAIPHEI, K., MAHMOOD, A. & MAHMOOD, S. 2015. Frequent activation of the beta-catenin gene in sporadic colorectal carcinomas: A mutational & expression analysis. *Mol Carcinog***,** 1627-1638.

ARANDA, P. S., LAJOIE, D. M. & JORCYK, C. L. 2012. Bleach gel: a simple agarose gel for analyzing RNA quality. *Electrophoresis,* 33**,** 366-9.

ARFÈ, A., MALVEZZI, M., BERTUCCIO, P., DECARLI, A., LA VECCHIA, C. & NEGRI, E. 2011. Cancer mortality trend analysis in Italy, 1970-2007. *Eur J Cancer Prev,* 20**,** 364-74.

ARNOLD, M., SIERRA, M. S., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2016. Global patterns and trends in colorectal cancer incidence and mortality. *Gut***,** 1-9.

ARORA, G., MANNALITHARA, A., SINGH, G., GERSON, L. B. & TRIADAFILOPOULOS, G. 2009. Risk of perforation from a colonoscopy in adults: a large population-based study. *Gastrointest Endosc,* 69**,** 654-64.

ARTEAGA, C. L. 2002. Epidermal growth factor receptor dependence in human tumors: more than just expression? *Oncologist,* 7 Suppl 4**,** 31-9.

ASSIS, R. V. B. F. 2011. Rastreamento e Vigilância do Câncer Colorretal: Guidelines Mundiais. *GED gastroenterol. endosc. dig,* 30**,** 62-74.

ATKIN, W. S., EDWARDS, R., KRALJ-HANS, I., WOOLDRAGE, K., HART, A. R., NORTHOVER, J. M., PARKIN, D. M., WARDLE, J., DUFFY, S. W., CUZICK, J. & INVESTIGATORS, U. K. F. S. T. 2010. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet,* 375**,** 1624-33.

ATKINSON, A. J., COLBURN, W. A., DEGRUTTOLA, V. G., DEMETS, D. L., DOWNING, G. J., HOTH, D. F., OATES, J. A., PECK, C. C., SCHOOLEY, R. T., SPILKER, B. A., WOODCOCK, J. & ZEGER, S. L. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics,* 69**,** 89-95.

AUDIGIER, Y., PICAULT, F. X., CHAVES-ALMAGRO, C. & MASRI, B. 2013. G Protein-Coupled Receptors in cancer: biochemical interactions and drug design. *Prog Mol Biol Transl Sci,* 115**,** 143-73.

AZEVEDO, M. R., HORTA, B. L., GIGANTE, D. P., VICTORA, C. G. & BARROS, F. C. 2008. [Factors associated to leisure-time sedentary lifestyle in adults of 1982 birth cohort, Pelotas, Southern Brazil]. *Rev Saude Publica,* 42 Suppl 2**,** 70-7.

BAARSMA, H. A., KONIGSHOFF, M. & GOSENS, R. 2013. The WNT signaling pathway from ligand secretion to gene transcription: molecular mechanisms and pharmacological targets. *Pharmacol Ther,* 138**,** 66-83.

BABA, Y., ISHIMOTO, T., KURASHIGE, J., IWATSUKI, M., SAKAMOTO, Y., YOSHIDA, N., WATANABE, M. & BABA, H. 2016. Mini-review: Epigenetic field cancerization in gastrointestinal cancers. *Cancer Letters,* 375**,** 360-366.

BAE, J. M., KIM, J. H. & KANG, G. H. 2013. Epigenetic alterations in colorectal cancer: the CpG island methylator phenotype. *Histol Histopathol,* 28**,** 585-95.

BAI, D., ZHANG, J., XIAO, W. & ZHENG, X. 2014. Regulation of the HDM2-p53 pathway by ribosomal protein L6 in response to ribosomal stress. *Nucleic Acids Res,* 42**,** 1799-811.

BAKER, S. J., MARKOWITZ, S., FEARON, E. R., WILLSON, J. K. & VOGELSTEIN, B. 1990a. Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science,* 249**,** 912-5.

BAKER, S. J., PREISINGER, A. C., JESSUP, J. M., PARASKEVA, C., MARKOWITZ, S., WILLSON, J. K., HAMILTON, S. & VOGELSTEIN, B. 1990b. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res,* 50**,** 7717-22.

BARKAN, D., GREEN, J. E. & CHAMBERS, A. F. 2010. Extracellular matrix: a gatekeeper in the transition from dormancy to metastatic growth. *Eur J Cancer,* 46**,** 1181-8.

BASS, B. P., ENGEL, K. B., GREYTAK, S. R. & MOORE, H. M. 2014. A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen? *Arch Pathol Lab Med,* 138**,** 1520-30.

BELL, H. S. & RYAN, K. M. 2005. Intracellular signalling and cancer: complex pathways lead to multiple targets. *Eur J Cancer,* 41**,** 206-15.

BENSON, A. B., 3RD, SCHRAG, D., SOMERFIELD, M. R., COHEN, A. M., FIGUEREDO, A. T., FLYNN, P. J., KRZYZANOWSKA, M. K., MAROUN, J., MCALLISTER, P., VAN CUTSEM, E., BROUWERS, M., CHARETTE, M. & HALLER, D. G. 2004. American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. *J Clin Oncol,* 22**,** 3408-19.

BENSON, V. S., PATNICK, J., DAVIES, A. K., NADEL, M. R., SMITH, R. A., ATKIN, W. S. & INTERNATIONAL COLORECTAL CANCER SCREENING, N. 2008. Colorectal cancer screening: a comparison of 35 initiatives in 17 countries. *Int J Cancer,* 122**,** 1357-67.

BENTON, S. C., SEAMAN, H. E. & HALLORAN, S. P. 2015. Faecal occult blood testing for colorectal cancer screening: the past or the future. *Curr Gastroenterol Rep,* 17**,** 428.

BERNAL, M., GÓMEZ, F. J. & GÓMEZ, G. 2009. Trends in cancer mortality in Spain: 1975-2004. *Tumori,* 95**,** 669-74.

BERNSTEIN, H. S. & COUGHLIN, S. R. 1998. A mammalian homolog of fission yeast Cdc5 regulates G2 progression and mitotic entry. *J Biol Chem,* 273**,** 4666-71.

BEROUKHIM, R., MERMEL, C. H., PORTER, D., WEI, G., RAYCHAUDHURI, S., DONOVAN, J., BARRETINA, J., BOEHM, J. S., DOBSON, J., URASHIMA, M., MC HENRY, K. T., PINCHBACK, R. M., LIGON, A. H., CHO, Y. J., HAERY, L., GREULICH, H., REICH, M., WINCKLER, W., LAWRENCE, M. S., WEIR, B. A., TANAKA, K. E., CHIANG, D. Y., BASS, A. J., LOO, A., HOFFMAN, C., PRENSNER, J., LIEFELD, T., GAO, Q., YECIES, D., SIGNORETTI, S., MAHER, E., KAYE, F. J., SASAKI, H., TEPPER, J. E., FLETCHER, J. A., TABERNERO, J., BASELGA, J., TSAO, M. S., DEMICHELIS, F., RUBIN, M. A., JANNE, P. A., DALY, M. J., NUCERA, C., LEVINE, R. L., EBERT, B. L., GABRIEL, S., RUSTGI, A. K., ANTONESCU, C. R., LADANYI, M., LETAI, A., GARRAWAY, L. A., LODA, M., BEER, D. G., TRUE, L. D., OKAMOTO, A., POMEROY, S. L., SINGER, S., GOLUB, T. R., LANDER, E. S., GETZ, G., SELLERS, W. R. & MEYERSON, M. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature,* 463**,** 899-905.

BHOWMICK, N. A., NEILSON, E. G. & MOSES, H. L. 2004. Stromal fibroblasts in cancer initiation and progression. *Nature,* 432**,** 332-7.

BIAMONTI, G., CATILLO, M., PIGNATARO, D., MONTECUCCO, A. & GHIGNA, C. 2014. The alternative splicing side of cancer. *Semin Cell Dev Biol,* 32**,** 30-6.

BIANCHI, P., LAGHI, L., DELCONTE, G. & MALESCI, A. 2011. Prognostic value of colorectal cancer biomarkers. *Cancers (Basel),* 3**,** 2080-105.

BLACKBURN, E. H. 1991. Structure and function of telomeres. *Nature,* 350**,** 569.

BLACKBURN, G. F., SHAH, H. P., KENTEN, J. H., LELAND, J., KAMIN, R. A., LINK, J., PETERMAN, J., POWELL, M. J., SHAH, A., TALLEY, D. B. & ET AL. 1991. Electrochemiluminescence detection for development of immunoassays and DNA probe assays for clinical diagnostics. *Clin Chem,* 37**,** 1534-9.

BOLAND, C. R., THIBODEAU, S. N., HAMILTON, S. R., SIDRANSKY, D., ESHLEMAN, J. R., BURT, R. W., MELTZER, S. J., RODRIGUEZ-BIGAS, M. A., FODDE, R., RANZANI, G. N. & SRIVASTAVA, S. 1998. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res,* 58**,** 5248-57.

BONDAR, T. & MEDZHITOV, R. 2013. The origins of tumor-promoting inflammation. *Cancer Cell,* 24**,** 143-4.

BOSCH, L. J., CARVALHO, B., FIJNEMAN, R. J., JIMENEZ, C. R., PINEDO, H. M., VAN ENGELAND, M. & MEIJER, G. A. 2011. Molecular tests for colorectal cancer screening. *Clin Colorectal Cancer,* 10**,** 8-23.

BRADBURN, M. J., CLARK, T. G., LOVE, S. B. & ALTMAN, D. G. 2003. Survival analysis Part III: multivariate data analysis -- choosing a model and assessing its adequacy and fit. *Br J Cancer,* 89**,** 605-11.

BRENNER, H., HOFFMEISTER, M., BRENNER, G., ALTENHOFEN, L. & HAUG, U. 2009. Expected reduction of colorectal cancer incidence within 8 years after introduction of the German screening colonoscopy programme: estimates based on 1,875,708 screening colonoscopies. *Eur J Cancer,* 45**,** 2027-33.

BRENNER, H., KLOOR, M. & POX, C. P. 2014. Colorectal cancer. *Lancet,* 383**,** 1490-502.

BRUUN, J., KOLBERG, M., NESLAND, J. M., SVINDLAND, A., NESBAKKEN, A. & LOTHE, R. A. 2014. Prognostic Significance of beta-Catenin, E-Cadherin, and SOX9 in Colorectal Cancer: Results from a Large Population-Representative Series. *Front Oncol,* 4**,** 118.

BUELL, J. F., GROSS, T. G. & WOODLE, E. S. 2005. Malignancy after transplantation. *Transplantation,* 80**,** S254-64.

BUIE, W. D. & ATTARD, J. A. P. 2005. Follow-Up Recommendations for Colon Cancer. *Clin Colon Rectal Surg,* 18**,** 232-43.

BUNZ, F., DUTRIAUX, A., LENGAUER, C., WALDMAN, T., ZHOU, S., BROWN, J. P., SEDIVY, J. M., KINZLER, K. W. & VOGELSTEIN, B. 1998. Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science,* 282**,** 1497-501.

BURKE, H. B. 2004. Outcome prediction and the future of the TNM staging system. *J Natl Cancer Inst,* 96**,** 1408-9.

BUSTIN, S., DHILLON, H. S., KIRVELL, S., GREENWOOD, C., PARKER, M., SHIPLEY, G. L. & NOLAN, T. 2015. Variability of the reverse transcription step: practical implications. *Clin Chem,* 61**,** 202-12.

BUSTIN, S. A., BENES, V., GARSON, J., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G., WITTWER, C. T., SCHJERLING, P., DAY, P. J., ABREU, M., AGUADO, B., BEAULIEU, J. F., BECKERS, A., BOGAERT, S., BROWNE, J. A., CARRASCO-RAMIRO, F., CEELEN, L., CIBOROWSKI, K., CORNILLIE, P., COULON, S., CUYPERS, A., DE BROUWER, S., DE CEUNINCK, L., DE CRAENE, J., DE NAEYER, H., DE SPIEGELAERE, W., DECKERS, K., DHEEDENE, A., DURINCK, K., FERREIRA-TEIXEIRA, M., FIEUW, A., GALLUP, J. M., GONZALO-FLORES, S., GOOSSENS, K., HEINDRYCKX, F., HERRING, E., HOENICKA, H., ICARDI, L., JAGGI, R., JAVAD, F., KARAMPELIAS, M., KIBENGE, F., KIBENGE, M., KUMPS, C., LAMBERTZ, I., LAMMENS, T., MARKEY, A., MESSIAEN, P., METS, E., MORAIS, S., MUDARRA-RUBIO, A., NAKIWALA, J., NELIS, H., OLSVIK, P. A., PEREZ-NOVO, C., PLUSQUIN, M., REMANS, T., RIHANI, A., RODRIGUES-SANTOS, P., RONDOU, P., SANDERS, R., SCHMIDT-BLEEK, K., SKOVGAARD, K., SMEETS, K., TABERA, L., TOEGEL, S., VAN ACKER, T., VAN DEN BROECK, W., VAN DER MEULEN, J., VAN GELE, M., VAN PEER, G., VAN POUCKE, M.,

VAN ROY, N., VERGULT, S., WAUMAN, J., TSHUIKINA-WIKLANDER, M., WILLEMS, E., ZACCARA, S., ZEKA, F. & VANDESOMPELE, J. 2013. The need for transparency and good practices in the qPCR literature. *Nat Methods,* 10**,** 1063-7.

BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J. & WITTWER, C. T. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* 55**,** 611-22.

CANCER GENOME ATLAS NETWORK 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature,* 487**,** 330-7.

CAO, Y., LIANG, H., ZHANG, F., LUAN, Z., ZHAO, S., WANG, X. A., LIU, S., BAO, R., SHU, Y., MA, Q., ZHU, J. & LIU, Y. 2016. Prohibitin overexpression predicts poor prognosis and promotes cell proliferation and invasion through ERK pathway activation in gallbladder cancer. *J Exp Clin Cancer Res,* 35**,** 68.

CARMELIET, P. 2005. VEGF as a key mediator of angiogenesis in cancer. *Oncology,* 69 Suppl 3**,** 4-10.

CENSO, I. 2010. Instituto Brasileiro de Geografia e Estatística.

CENTER, M. M., JEMAL, A. & WARD, E. 2009. International trends in colorectal cancer incidence rates. *Cancer Epidemiol Biomarkers Prev,* 18**,** 1688-94.

CHAMBERLAIN, B. K., FADARE, O. & DESOUKI, M. M. 2015. The role of immunohistochemistry in the evaluation of gynecologic pathology: a single institutional experience. *Ann Diagn Pathol,* 19**,** 88-90.

CHAN, S. K., GRIFFITH, O. L., TAI, I. T. & JONES, S. J. 2008. Meta-analysis of colorectal cancer gene expression profiling studies identifies consistently reported candidate biomarkers. *Cancer Epidemiol Biomarkers Prev,* 17**,** 543-52.

CHANDRAN, U. R., DHIR, R., MA, C., MICHALOPOULOS, G., BECICH, M. & GILBERTSON, J. 2005. Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. *BMC Cancer,* 5**,** 45.

CHATENOUD, L., BERTUCCIO, P., BOSETTI, C., LEVI, F., CURADO, M. P., MALVEZZI, M., NEGRI, E. & LA VECCHIA, C. 2010. Trends in cancer mortality in Brazil, 1980-2004. *Eur J Cancer Prev,* 19**,** 79-86.

CHATURVEDI, S., SIEGEL, D., WAGNER, C. L., PARK, J., VAN DE VELDE, H., VERMEULEN, J., FUNG, M. C., REDDY, M., HALL, B. & SASSER, K. 2015. Development and validation of panoptic Meso scale discovery assay to quantify total systemic interleukin-6. *Br J Clin Pharmacol,* 80**,** 687-97.

CHAVEZ-PEREZ, V. A., STRASBERG-RIEBER, M. & RIEBER, M. 2014. Hypoxia and hypoxia mimetic cooperate to counteract tumor cell resistance to glucose starvation preferentially in tumor cells with mutant p53. *Biochem Biophys Res Commun,* 443**,** 120-5.

CHEN, D., CHEN, F., LU, X., YANG, X., XU, Z., PAN, J., HUANG, Y., LIN, H. & CHI, P. 2010a. Identification of prohibitin as a potential biomarker for colorectal carcinoma based on proteomics technology. *Int J Oncol,* 37**,** 355-65.

CHEN, Y., QIU, Z., KAMRUZZAMAN, A., SNODGRASS, T., SCARFE, A. & BRYANT, H. E. 2010b. Survival of metastatic colorectal cancer patients treated with chemotherapy in Alberta (1995-2004). *Support Care Cancer,* 18**,** 217-24.

CHEN, Z., HE, X., JIA, M., LIU, Y., QU, D., WU, D., WU, P., NI, C., ZHANG, Z., YE, J., XU, J. & HUANG, J. 2013. β-catenin overexpression in the nucleus predicts progress disease and unfavourable survival in colorectal cancer: a meta-analysis. *PLoS One,* 8**,** e63854.

CHEVREUL, K. 2010. Colorectal cancer in France. *Eur J Health Econ,* 10 Suppl 1**,** S15-20.

CHOWDHURY, D., KUMAR, D., BHADRA, U., DEVI, T. A. & BHADRA, M. P. 2017. Prohibitin confers cytoprotection against ISO-induced hypertrophy in H9c2 cells via

attenuation of oxidative stress and modulation of Akt/Gsk-3β signaling. *Mol Cell Biochem,* 425**,** 155-168.
CHOWDHURY, I., THOMPSON, W. E. & THOMAS, K. 2014. Prohibitins role in cellular survival through Ras-Raf-MEK-ERK pathway. *J Cell Physiol,* 229**,** 998-1004.
CLARK, J. C., COLLAN, Y., EIDE, T. J., ESTÈVE, J., EWEN, S., GIBBS, N. M., JENSEN, O. M., KOSKELA, E., MACLENNAN, R. & SIMPSON, J. G. 1985. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. *Int J Cancer,* 36**,** 179-86.
CLARK-LANGONE, K. M., SANGLI, C., KRISHNAKUMAR, J. & WATSON, D. 2010. Translating tumor biology into personalized treatment planning: analytical performance characteristics of the Oncotype DX Colon Cancer Assay. *BMC Cancer,* 10**,** 691.
CLARK-LANGONE, K. M., WU, J. Y., SANGLI, C., CHEN, A., SNABLE, J. L., NGUYEN, A., HACKETT, J. R., BAKER, J., YOTHERS, G., KIM, C. & CRONIN, M. T. 2007. Biomarker discovery for colon cancer using a 761 gene RT-PCR assay. *BMC Genomics,* 8**,** 279.
CLASSON, M. & HARLOW, E. 2002. The retinoblastoma tumour suppressor in development and cancer. *Nat Rev Cancer,* 2**,** 910-7.
CLEVERS, H. & NUSSE, R. 2012. Wnt/β-catenin signaling and disease. *Cell,* 149**,** 1192-205.
COGHLIN, C. & MURRAY, G. I. 2015. Biomarkers of colorectal cancer: recent advances and future challenges. *Proteomics Clin Appl,* 9**,** 64-71.
COLLER, H. A. 2007. What's taking so long? S-phase entry from quiescence versus proliferation. *Nat Rev Mol Cell Biol,* 8**,** 667-70.
COLLINS, K. & MITCHELL, J. R. 2002. Telomerase in the human organism. *Oncogene,* 21**,** 564-79.
CONNER, J. R. & HORNICK, J. L. 2015. Metastatic carcinoma of unknown primary: diagnostic approach using immunohistochemistry. *Adv Anat Pathol,* 22**,** 149-67.
CONNETT, J. E. & LEE, W. W. 1990. Estimation of the coefficient of variation from laboratory analysis of split specimens for quality control in clinical trials. *Control Clin Trials,* 11**,** 24-36.
CORY, S. & ADAMS, J. M. 2005. Killing cancer cells by flipping the Bcl-2/Bax switch. *Cancer Cell,* 8**,** 5-6.
COUSSENS, L. M. & WERB, Z. 2002. Inflammation and cancer. *Nature,* 420**,** 860-7.
CRISTOFALO, V. J. & PIGNOLO, R. J. 1993. Replicative senescence of human fibroblast-like cells in culture. *Physiol Rev,* 73**,** 617-38.
CROCE, C. M. 2008. Oncogenes and cancer. *N Engl J Med,* 358**,** 502-11.
CROCETTI, E., BUZZONI, C. & ZAPPA, M. 2010. Colorectal cancer incidence rates have decreased in central Italy. *Eur J Cancer Prev,* 19**,** 424-5.
CROCETTI, E., DE ANGELIS, R., BUZZONI, C., MARIOTTO, A., STORM, H., COLONNA, M., ZANETTI, R., SERRAINO, D., MICHIARA, M., CIRILLI, C., IANNELLI, A., MAZZOLENI, G., SECHI, O., SANOJA GONZALEZ, M. E., GUZZINATI, S., CAPOCACCIA, R. & DAL MASO, L. 2013. Cancer prevalence in United States, Nordic Countries, Italy, Australia, and France: an analysis of geographic variability. *Br J Cancer,* 109**,** 219-28.
CUELLO-CARRION, F. D., SHORTREDE, J. E., ALVAREZ-OLMEDO, D., CAYADO-GUTIERREZ, N., CASTRO, G. N., ZOPPINO, F. C., GUERRERO, M., MARTINIS, E., WUILLOUD, R., GOMEZ, N. N., BIAGGIO, V., OROZCO, J., GAGO, F. E., CIOCCA, L. A., FANELLI, M. A. & CIOCCA, D. R. 2015. HER2 and beta-catenin protein location: importance in the prognosis of breast cancer patients and their correlation when breast cancer cells suffer stressful situations. *Clin Exp Metastasis,* 32**,** 151-68.
CZABOTAR, P. E., LESSENE, G., STRASSER, A. & ADAMS, J. M. 2014. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat Rev Mol Cell Biol,* 15**,** 49-63.
DANG, C. V. 2012. MYC on the path to cancer. *Cell,* 149**,** 22-35.

DAS NEVES, F. J., MATTOS, I. E. & KOIFMAN, R. J. 2005. [Colon and rectal cancer mortality in Brazilian capitals, 1980-1997]. *Arq Gastroenterol,* 42**,** 63-70.

DATTATREYA, S. 2013. Metastatic colorectal cancer-prolonging overall survival with targeted therapies. *South Asian J Cancer,* 2**,** 179-85.

DE ROOCK, W., CLAES, B., BERNASCONI, D., DE SCHUTTER, J., BIESMANS, B., FOUNTZILAS, G., KALOGERAS, K. T., KOTOULA, V., PAPAMICHAEL, D., LAURENT-PUIG, P., PENAULT-LLORCA, F., ROUGIER, P., VINCENZI, B., SANTINI, D., TONINI, G., CAPPUZZO, F., FRATTINI, M., MOLINARI, F., SALETTI, P., DE DOSSO, S., MARTINI, M., BARDELLI, A., SIENA, S., SARTORE-BIANCHI, A., TABERNERO, J., MACARULLA, T., DI FIORE, F., GANGLOFF, A. O., CIARDIELLO, F., PFEIFFER, P., QVORTRUP, C., HANSEN, T. P., VAN CUTSEM, E., PIESSEVAUX, H., LAMBRECHTS, D., DELORENZI, M. & TEJPAR, S. 2010. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol,* 11**,** 753-62.

DE WIT, M., FIJNEMAN, R. J., VERHEUL, H. M., MEIJER, G. A. & JIMENEZ, C. R. 2013. Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. *Clin Biochem,* 46**,** 466-79.

DE WIT, M., KANT, H., PIERSMA, S. R., PHAM, T. V., MONGERA, S., VAN BERKEL, M. P., BOVEN, E., PONTÉN, F., MEIJER, G. A., JIMENEZ, C. R. & FIJNEMAN, R. J. 2014. Colorectal cancer candidate biomarkers identified by tissue secretome proteome profiling. *J Proteomics,* 99**,** 26-39.

DEJARDIN, O., RACHET, B., MORRIS, E., BOUVIER, V., JOOSTE, V., HAYNES, R., COOMBES, E. G., FORMAN, D., JONES, A. P., BOUVIER, A. M. & LAUNOY, G. 2013. Management of colorectal cancer explains differences in 1-year relative survival between France and England for patients diagnosed 1997-2004. *Br J Cancer,* 108**,** 775-83.

DEL CAMPO, M., JONGBLOED, W., TWAALFHOVEN, H. A., VEERHUIS, R., BLANKENSTEIN, M. A. & TEUNISSEN, C. E. 2015. Facilitating the Validation of Novel Protein Biomarkers for Dementia: An Optimal Workflow for the Development of Sandwich Immunoassays. *Front Neurol,* 6**,** 202.

DEL CASAR LIZCANO, J. M., VIZOSO PINEIRO, F., GONZALEZ SANCHEZ, L. O., MARTIN SUAREZ, A., GAVA, R., CUESTA FERNANDEZ, E. & DIEZ SANTISTEBAN, M. C. 2003. [Expression and clinical significance of collagenase-3 (MMP-13) in gastric cancer]. *Gastroenterol Hepatol,* 26**,** 1-7.

DENKO, N. C. 2008. Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nat Rev Cancer,* 8**,** 705-13.

DEVOS, T., TETZNER, R., MODEL, F., WEISS, G., SCHUSTER, M., DISTLER, J., STEIGER, K. V., GRUTZMANN, R., PILARSKY, C., HABERMANN, J. K., FLESHNER, P. R., OUBRE, B. M., DAY, R., SLEDZIEWSKI, A. Z. & LOFTON-DAY, C. 2009. Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. *Clin Chem,* 55**,** 1337-46.

DEWHURST, C., ROSEN, M. P., BLAKE, M. A., BAKER, M. E., CASH, B. D., FIDLER, J. L., GREENE, F. L., HINDMAN, N. M., JONES, B., KATZ, D. S., LALANI, T., MILLER, F. H., SMALL, W. C., SUDAKOFF, G. S., TULCHINSKY, M., YAGHMAI, V. & YEE, J. 2012. ACR Appropriateness Criteria pretreatment staging of colorectal cancer. *J Am Coll Radiol,* 9**,** 775-81.

DI LENA, M., TRAVAGLIO, E. & ALTOMARE, D. F. 2013. New strategies for colorectal cancer screening. *World J Gastroenterol,* 19**,** 1855-60.

DIAMANDIS, E. P. 2014. Present and future of cancer biomarkers. *Clin Chem Lab Med,* 52**,** 791-4.

DICK, F. A. & RUBIN, S. M. 2013. Molecular mechanisms underlying RB protein function. *Nat Rev Mol Cell Biol,* 14**,** 297-306.

DIENSTMANN, R., VERMEULEN, L., GUINNEY, J., KOPETZ, S., TEJPAR, S. & TABERNERO, J. 2017. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer,* 17**,** 268.

DIJKSTRA, J. R., VAN KEMPEN, L. C., NAGTEGAAL, I. D. & BUSTIN, S. A. 2014. Critical appraisal of quantitative PCR results in colorectal cancer research: can we rely on published qPCR results? *Mol Oncol,* 8**,** 813-8.

DILLER, L., KASSEL, J., NELSON, C. E., GRYKA, M. A., LITWAK, G., GEBHARDT, M., BRESSAC, B., OZTURK, M., BAKER, S. J. & VOGELSTEIN, B. 1990. p53 functions as a cell cycle control protein in osteosarcomas. *Mol Cell Biol,* 10**,** 5772-81.

DREW, D. A., CAO, Y. & CHAN, A. T. 2016. Aspirin and colorectal cancer: the promise of precision chemoprevention. *Nat Rev Cancer,* 16**,** 173-86.

DROZDOV, I., KIDD, M., NADLER, B., CAMP, R. L., MANE, S. M., HAUSO, O., GUSTAFSSON, B. I. & MODLIN, I. M. 2009. Predicting neuroendocrine tumor (carcinoid) neoplasia using gene expression profiling and supervised machine learning. *Cancer,* 115**,** 1638-50.

DU, J., SHI, Y., PAN, Y., JIN, X., LIU, C., LIU, N., HAN, Q., LU, Y., QIAO, T. & FAN, D. 2005. Regulation of multidrug resistance by ribosomal protein l6 in gastric cancer cells. *Cancer Biol Ther,* 4**,** 242-7.

DUESBERG, P. & LI, R. 2003. Multistep carcinogenesis: a chain reaction of aneuploidizations. *Cell Cycle,* 2**,** 202-10.

DUFFY, M. J., VAN ROSSUM, L. G., VAN TURENHOUT, S. T., MALMINIEMI, O., STURGEON, C., LAMERZ, R., NICOLINI, A., HAGLUND, C., HOLUBEC, L., FRASER, C. G. & HALLORAN, S. P. 2011. Use of faecal markers in screening for colorectal neoplasia: a European group on tumor markers position paper. *Int J Cancer,* 128**,** 3-11.

DUMITH, S. C., HALLAL, P. C., REIS, R. S. & KOHL, H. W., 3RD 2011. Worldwide prevalence of physical inactivity and its association with human development index in 76 countries. *Prev Med,* 53**,** 24-8.

DUNN, G. P., BRUCE, A. T., IKEDA, H., OLD, L. J. & SCHREIBER, R. D. 2002. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol,* 3**,** 991-8.

DUNN, G. P., OLD, L. J. & SCHREIBER, R. D. 2004. The three Es of cancer immunoediting. *Annu Rev Immunol,* 22**,** 329-60.

EDMOND, V., MOYSAN, E., KHOCHBIN, S., MATTHIAS, P., BRAMBILLA, C., BRAMBILLA, E., GAZZERI, S. & EYMIN, B. 2011. Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. *EMBO J,* 30**,** 510-23.

EDWARDS, B. K., WARD, E., KOHLER, B. A., EHEMAN, C., ZAUBER, A. G., ANDERSON, R. N., JEMAL, A., SCHYMURA, M. J., LANSDORP-VOGELAAR, I., SEEFF, L. C., VAN BALLEGOOIJEN, M., GOEDE, S. L. & RIES, L. A. 2010. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer,* 116**,** 544-73.

ELMORE, S. 2007. Apoptosis: a review of programmed cell death. *Toxicol Pathol,* 35**,** 495-516.

ELMUNZER, B. J., HAYWARD, R. A., SCHOENFELD, P. S., SAINI, S. D., DESHPANDE, A. & WALJEE, A. K. 2012. Effect of flexible sigmoidoscopy-based screening on incidence and mortality of colorectal cancer: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med,* 9**,** e1001352.

ENGEMANN, H., HEINZEL, V., PAGE, G., PREUSS, U. & SCHEIDTMANN, K. H. 2002. DAP-like kinase interacts with the rat homolog of Schizosaccharomyces pombe CDC5 protein, a factor involved in pre-mRNA splicing and required for G2/M phase transition. *Nucleic Acids Res,* 30**,** 1408-17.

ENGVALL, E. & PERLMANN, P. 1971. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry,* 8**,** 871-4.

ERSTAD, D. J., TUMUSIIME, G. & CUSACK, J. C., JR. 2015. Prognostic and Predictive Biomarkers in Colorectal Cancer: Implications for the Clinical Surgeon. *Ann Surg Oncol,* 22**,** 3433-50.

FAN, W., YANG, H., LIU, T., WANG, J., LI, T. W., MAVILA, N., TANG, Y., YANG, J., PENG, H., TU, J., ANNAMALAI, A., NOUREDDIN, M., KRISHNAN, A., GORES, G. J., MARTÍNEZ-CHANTAR, M. L., MATO, J. M. & LU, S. C. 2016. Prohibitin 1 suppresses liver cancer tumorigenesis in mice and human hepatocellular and cholangiocarcinoma cells. *Hepatology*.

FEARON, E. R. 2011. Molecular genetics of colorectal cancer. *Annu Rev Pathol,* 6**,** 479-507.

FEARON, E. R. & VOGELSTEIN, B. 1990. A genetic model for colorectal tumorigenesis. *Cell,* 61**,** 759-67.

FERLAY, J., SOERJOMATARAM, I., DIKSHIT, R., ESER, S., MATHERS, C., REBELO, M., PARKIN, D. M., FORMAN, D. & BRAY, F. 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer,* 136**,** E359-86.

FERRARA, N. 2010. Pathways mediating VEGF-independent tumor angiogenesis. *Cytokine Growth Factor Rev,* 21**,** 21-6.

FERRI, C. P. 2012. Population ageing in Latin America: dementia and related disorders. *Rev Bras Psiquiatr,* 34**,** 371-4.

FIDLER, I. J. 1970. Metastasis: guantitative analysis of distribution and fate of tumor embolilabeled with 125 I-5-iodo-2'-deoxyuridine. *J Natl Cancer Inst,* 45**,** 773-82.

FIGUEIROA BDE, Q., VANDERLEI, L. C., FRIAS, P. G., CARVALHO, P. I. & SZWARCWALD, C. L. 2013. [Analysis of coverage in the Mortality Information System in Olinda, Pernambuco State, Brazil]. *Cad Saude Publica,* 29**,** 475-84.

FIRE, A., XU, S., MONTGOMERY, M. K., KOSTAS, S. A., DRIVER, S. E. & MELLO, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature,* 391**,** 806-11.

FRANCA, E., DE ABREU, D. X., RAO, C. & LOPEZ, A. D. 2008. Evaluation of cause-of-death statistics for Brazil, 2002-2004. *Int J Epidemiol,* 37**,** 891-901.

FRANZONI, A., DIMA, M., D'AGOSTINO, M., PUPPIN, C., FABBRO, D., LORETO, C. D., PANDOLFI, M., PUXEDDU, E., MORETTI, S., CELANO, M., BRUNO, R., FILETTI, S., RUSSO, D. & DAMANTE, G. 2009. Prohibitin is overexpressed in papillary thyroid carcinomas bearing the BRAF(V600E) mutation. *Thyroid,* 19**,** 247-55.

FRIDMAN, J. S. & LOWE, S. W. 2003. Control of apoptosis by p53. *Oncogene,* 22**,** 9030-40.

FRIEDL, P. & ALEXANDER, S. 2011. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell,* 147**,** 992-1009.

FRIEND, S. H., BERNARDS, R., ROGELJ, S., WEINBERG, R. A., RAPAPORT, J. M., ALBERT, D. M. & DRYJA, T. P. 1986. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature,* 323**,** 643-6.

FU, X. D. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA,* 1**,** 663-80.

FUNAIOLI, C., LONGOBARDI, C. & MARTONI, A. A. 2008. The impact of chemotherapy on overall survival and quality of life of patients with metastatic colorectal cancer: a review of phase III trials. *J Chemother,* 20**,** 14-27.

FUNG, K. Y., TABOR, B., BUCKLEY, M. J., PRIEBE, I. K., PURINS, L., POMPEIA, C., BRIERLEY, G. V., LOCKETT, T., GIBBS, P., TIE, J., MCMURRICK, P., MOORE, J., RUSZKIEWICZ, A., NICE, E., ADAMS, T. E., BURGESS, A. & COSGROVE, L. J. 2015. Blood-based protein biomarker panel for the detection of colorectal cancer. *PLoS One,* 10**,** e0120425.

GALLUP, J. M. & ACKERMANN, M. R. 2006. Addressing fluorogenic real-time qPCR inhibition using the novel custom Excel file system 'FocusField2-6GallupqPCRSet-upTool-001' to attain consistently high fidelity qPCR reactions. *Biol Proced Online,* 8**,** 87-152.

GALLUZZI, L., BRAVO-SAN PEDRO, J. M. & KROEMER, G. 2017. Mitophagy: Permitted by Prohibitin. *Curr Biol,* 27**,** R73-R76.

GAO, K. & HUANG, L. 2009. Non-viral Methods for siRNA Delivery. *Molecular pharmaceutics,* 6**,** 651-658.
GEIERSBACH, K. B. & SAMOWITZ, W. S. 2011. Microsatellite instability and colorectal cancer. *Arch Pathol Lab Med,* 135**,** 1269-77.
GEKID. 2013. *Atlas of cancer incidence and mortality of the Association of Population-based Cancer Registries in Germany* [Online]. Association of Population-based Cancer Registries in Germany (GEKID). Available: http://www.ekr.med.uni-erlangen.de/GEKID/Atlas/CurrentVersion/Inzidenz/atlas.html [Accessed 25 Nov 2013].
GIALELI, C., THEOCHARIS, A. D. & KARAMANOS, N. K. 2011. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *Febs j,* 278**,** 16-27.
GIESSEN, C., LAUBENDER, R. P., ANKERST, D. P., STINTZING, S., MODEST, D. P., MANSMANN, U. & HEINEMANN, V. 2013. Progression-free survival as a surrogate endpoint for median overall survival in metastatic colorectal cancer: literature-based analysis from 50 randomized first-line trials. *Clin Cancer Res,* 19**,** 225-35.
GIOLO, S. R., SOLER, J. M., GREENWAY, S. C., ALMEIDA, M. A., DE ANDRADE, M., SEIDMAN, J. G., SEIDMAN, C. E., KRIEGER, J. E. & PEREIRA, A. C. 2012. Brazilian urban population genetic structure reveals a high degree of admixture. *Eur J Hum Genet,* 20**,** 111-6.
GLIMELIUS, B. 2013. Neo-adjuvant radiotherapy in rectal cancer. *World J Gastroenterol,* 19**,** 8489-501.
GONG, H., YAN, Y., FANG, B., XUE, Y., YIN, P., LI, L., ZHANG, G., SUN, X., CHEN, Z., MA, H., YANG, C., DING, Y., YONG, Y., ZHU, Y., YANG, H., KOMURO, I., GE, J. & ZOU, Y. 2014. Knockdown of nucleosome assembly protein 1-like 1 induces mesoderm formation and cardiomyogenesis via notch signaling in murine-induced pluripotent stem cells. *Stem Cells,* 32**,** 1759-73.
GONZALEZ-PONS, M. & CRUZ-CORREA, M. 2015. Colorectal Cancer Biomarkers: Where Are We Now? *Biomed Res Int,* 2015**,** 149014.
GOU, Y., SHI, Y., ZHANG, Y., NIE, Y., WANG, J., SONG, J., JIN, H., HE, L., GAO, L., QIAO, L., WU, K. & FAN, D. 2010. Ribosomal protein L6 promotes growth and cell cycle progression through upregulating cyclin E in gastric cancer cells. *Biochem Biophys Res Commun,* 393**,** 788-93.
GRADY, W. M., MYEROFF, L. L., SWINLER, S. E., RAJPUT, A., THIAGALINGAM, S., LUTTERBAUGH, J. D., NEUMANN, A., BRATTAIN, M. G., CHANG, J., KIM, S. J., KINZLER, K. W., VOGELSTEIN, B., WILLSON, J. K. & MARKOWITZ, S. 1999. Mutational inactivation of transforming growth factor beta receptor type II in microsatellite stable colon cancers. *Cancer Res,* 59**,** 320-4.
GRASER, A., STIEBER, P., NAGEL, D., SCHAFER, C., HORST, D., BECKER, C. R., NIKOLAOU, K., LOTTES, A., GEISBUSCH, S., KRAMER, H., WAGNER, A. C., DIEPOLDER, H., SCHIRRA, J., ROTH, H. J., SEIDEL, D., GOKE, B., REISER, M. F. & KOLLIGS, F. T. 2009. Comparison of CT colonography, colonoscopy, sigmoidoscopy and faecal occult blood tests for the detection of advanced adenoma in an average risk population. *Gut,* 58**,** 241-8.
GRAVELEY, B. R. 2000. Sorting out the complexity of SR protein functions. *RNA,* 6**,** 1197-211.
GREGORY, M. A. & HANN, S. R. 2000. c-Myc proteolysis by the ubiquitin-proteasome pathway: stabilization of c-Myc in Burkitt's lymphoma cells. *Mol Cell Biol,* 20**,** 2423-35.
GRILLO, F., PIGOZZI, S., CERIOLO, P., CALAMARO, P., FIOCCA, R. & MASTRACCI, L. 2015. Factors affecting immunoreactivity in long-term storage of formalin-fixed paraffin-embedded tissue sections. *Histochem Cell Biol,* 144**,** 93-9.

GRIVENNIKOV, S. I. & KARIN, M. 2011. Inflammatory cytokines in cancer: tumour necrosis factor and interleukin 6 take the stage. *Ann Rheum Dis,* 70 Suppl 1**,** i104-8.

GRUMOLATO, L., LIU, G., MONG, P., MUDBHARY, R., BISWAS, R., ARROYAVE, R., VIJAYAKUMAR, S., ECONOMIDES, A. N. & AARONSON, S. A. 2010. Canonical and noncanonical Wnts use a common mechanism to activate completely unrelated coreceptors. *Genes Dev,* 24**,** 2517-30.

GRUTZMANN, R., MOLNAR, B., PILARSKY, C., HABERMANN, J. K., SCHLAG, P. M., SAEGER, H. D., MIEHLKE, S., STOLZ, T., MODEL, F., ROBLICK, U. J., BRUCH, H. P., KOCH, R., LIEBENBERG, V., DEVOS, T., SONG, X., DAY, R. H., SLEDZIEWSKI, A. Z. & LOFTON-DAY, C. 2008. Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. *PLoS One,* 3**,** e3759.

GRÄUB, R., LANCERO, H., PEDERSEN, A., CHU, M., PADMANABHAN, K., XU, X. Q., SPITZ, P., CHALKLEY, R., BURLINGAME, A. L., STOKOE, D. & BERNSTEIN, H. S. 2008. Cell cycle-dependent phosphorylation of human CDC5 regulates RNA processing. *Cell Cycle,* 7**,** 1795-803.

GUGLIELMO-VIRET, V. & THULLIER, P. 2007. Comparison of an electrochemiluminescence assay in plate format over a colorimetric ELISA, for the detection of ricin B chain (RCA-B). *J Immunol Methods,* 328**,** 70-8.

GUIMARÃES, R. M., ROCHA, P. G., MUZI, C. D. & RAMOS, R. E. S. 2013. Increase income and mortality of colorrectal cancer in Brazil, 2001-2009. *Arq Gastroenterol,* 50**,** 64-9.

GUINNEY, J., DIENSTMANN, R., WANG, X., DE REYNIES, A., SCHLICKER, A., SONESON, C., MARISA, L., ROEPMAN, P., NYAMUNDANDA, G., ANGELINO, P., BOT, B. M., MORRIS, J. S., SIMON, I. M., GERSTER, S., FESSLER, E., DE SOUSA, E. M. F., MISSIAGLIA, E., RAMAY, H., BARRAS, D., HOMICSKO, K., MARU, D., MANYAM, G. C., BROOM, B., BOIGE, V., PEREZ-VILLAMIL, B., LADERAS, T., SALAZAR, R., GRAY, J. W., HANAHAN, D., TABERNERO, J., BERNARDS, R., FRIEND, S. H., LAURENT-PUIG, P., MEDEMA, J. P., SADANANDAM, A., WESSELS, L., DELORENZI, M., KOPETZ, S., VERMEULEN, L. & TEJPAR, S. 2015. The consensus molecular subtypes of colorectal cancer. *Nat Med,* 21**,** 1350-6.

GUO, F., HIROSHIMA, K., WU, D., SATOH, M., ABULAZI, M., YOSHINO, I., TOMONAGA, T., NOMURA, F. & NAKATANI, Y. 2012. Prohibitin in squamous cell carcinoma of the lung: its expression and possible clinical significance. *Hum Pathol,* 43**,** 1282-8.

GUPTA, S., SUSSMAN, D. A., DOUBENI, C. A., ANDERSON, D. S., DAY, L., DESHPANDE, A. R., ELMUNZER, B. J., LAIYEMO, A. O., MENDEZ, J., SOMSOUK, M., ALLISON, J., BHUKET, T., GENG, Z., GREEN, B. B., ITZKOWITZ, S. H. & MARTINEZ, M. E. 2014. Challenges and possible solutions to colorectal cancer screening for the underserved. *J Natl Cancer Inst,* 106**,** dju032.

GÜNES, C. & RUDOLPH, K. L. 2013. The role of telomeres in stem cells and cancer. *Cell,* 152**,** 390-3.

HAHN, W. C. & WEINBERG, R. A. 2002. Modelling the molecular circuitry of cancer. *Nat Rev Cancer,* 2**,** 331-41.

HAMMES, L. S., KORTE, J. E., TEKMAL, R. R., NAUD, P., EDELWEISS, M. I., VALENTE, P. T., LONGATTO-FILHO, A., KIRMA, N. & CUNHA-FILHO, J. S. 2007. Computer-assisted immunohistochemical analysis of cervical cancer biomarkers using low-cost and simple software. *Appl Immunohistochem Mol Morphol,* 15**,** 456-62.

HAMMOUDI, A., SONG, F., REED, K. R., JENKINS, R. E., MENIEL, V. S., WATSON, A. J., PRITCHARD, D. M., CLARKE, A. R. & JENKINS, J. R. 2013. Proteomic profiling of a mouse model of acute intestinal Apc deletion leads to identification of potential novel biomarkers of human colorectal cancer (CRC). *Biochem Biophys Res Commun,* 440**,** 364-70.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HANSEN, M., HANSEN, M. F. & CAVENEE, W. K. 1988. Tumor suppressors: Recessive mutations that lead to cancer. *Cell,* 53**,** 172.

HARDCASTLE, J. D., CHAMBERLAIN, J. O., ROBINSON, M. H., MOSS, S. M., AMAR, S. S., BALFOUR, T. W., JAMES, P. D. & MANGHAM, C. M. 1996. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet,* 348**,** 1472-7.

HARVEY, J. M., CLARK, G. M., OSBORNE, C. K. & ALLRED, D. C. 1999. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol,* 17**,** 1474-81.

HAUG, U., KUNTZ, K. M., KNUDSEN, A. B., HUNDT, S. & BRENNER, H. 2011. Sensitivity of immunochemical faecal occult blood testing for detecting left- vs right-sided colorectal neoplasia. *Br J Cancer,* 104**,** 1779-85.

HAYFLICK, L. & MOORHEAD, P. S. 1961. The serial cultivation of human diploid cell strains. *Exp Cell Res,* 25**,** 585-621.

HE, J., COOPER, H. M., REYES, A., DI RE, M., SEMBONGI, H., LITWIN, T. R., GAO, J., NEUMAN, K. C., FEARNLEY, I. M., SPINAZZOLA, A., WALKER, J. E. & HOLT, I. J. 2012. Mitochondrial nucleoid interacting proteins support mitochondrial protein synthesis. *Nucleic Acids Res,* 40**,** 6109-21.

HE, T. C., SPARKS, A. B., RAGO, C., HERMEKING, H., ZAWEL, L., DA COSTA, L. T., MORIN, P. J., VOGELSTEIN, B. & KINZLER, K. W. 1998. Identification of c-MYC as a target of the APC pathway. *Science,* 281**,** 1509-12.

HELPS, S. C., THORNTON, E., KLEINIG, T. J., MANAVIS, J. & VINK, R. 2012. Automatic nonsubjective estimation of antigen content visualized by immunohistochemistry using color deconvolution. *Appl Immunohistochem Mol Morphol,* 20**,** 82-90.

HERSZENYI, L., HRITZ, I., LAKATOS, G., VARGA, M. Z. & TULASSAY, Z. 2012. The behavior of matrix metalloproteinases and their inhibitors in colorectal cancer. *Int J Mol Sci,* 13**,** 13240-63.

HEWITSON, P., GLASZIOU, P., WATSON, E., TOWLER, B. & IRWIG, L. 2008. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. *Am J Gastroenterol,* 103**,** 1541-9.

HIGGINS, R. A., BLANKENSHIP, J. E. & KINNEY, M. C. 2008. Application of immunohistochemistry in the diagnosis of non-Hodgkin and Hodgkin lymphoma. *Arch Pathol Lab Med,* 132**,** 441-61.

HIGUCHI, R., DOLLINGER, G., WALSH, P. S. & GRIFFITH, R. 1992. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology (N Y),* 10**,** 413-7.

HOFF, G., GROTMOL, T., SKOVLUND, E. & BRETTHAUER, M. 2009. Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. *Bmj,* 338**,** b1846.

HOFMANN, H. S., HANSEN, G., RICHTER, G., TAEGE, C., SIMM, A., SILBER, R. E. & BURDACH, S. 2005. Matrix metalloproteinase-12 expression correlates with local recurrence and metastatic disease in non-small cell lung cancer patients. *Clin Cancer Res,* 11**,** 1086-92.

HOL, L., VAN LEERDAM, M. E., VAN BALLEGOOIJEN, M., VAN VUUREN, A. J., VAN DEKKEN, H., REIJERINK, J. C., VAN DER TOGT, A. C., HABBEMA, J. D. & KUIPERS, E. J. 2010. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. *Gut,* 59**,** 62-8.

HOLLAND, J. D., KLAUS, A., GARRATT, A. N. & BIRCHMEIER, W. 2013. Wnt signaling in stem and cancer stem cells. *Curr Opin Cell Biol,* 25**,** 254-64.

HORNICK, J. L. 2014. Novel uses of immunohistochemistry in the diagnosis and classification of soft tissue tumors. *Mod Pathol,* 27 Suppl 1**,** S47-63.

HSU, P. P. & SABATINI, D. M. 2008. Cancer cell metabolism: Warburg and beyond. *Cell,* 134**,** 703-7.

HUANG, S. H., LAW, C. H., KUO, P. H., HU, R. Y., YANG, C. C., CHUNG, T. W., LI, J. M., LIN, L. H., LIU, Y. C., LIAO, E. C., TSAI, Y. T., WEI, Y. S., LIN, C. C., CHANG, C. W., CHOU, H. C., WANG, W. C., CHANG, M. D., WANG, L. H., KUNG, H. J., CHAN, H. L. & LYU, P. C. 2016. MMP-13 is involved in oral cancer cell metastasis. *Oncotarget,* 7**,** 17144-61.

HUANG, Z., HUANG, D., NI, S., PENG, Z., SHENG, W. & DU, X. 2010. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. *Int J Cancer,* 127**,** 118-26.

HUNG, M. C. & LAU, Y. K. 1999. Basic science of HER-2/neu: a review. *Semin Oncol,* 26**,** 51-9.

HYNES, N. E. & MACDONALD, G. 2009. ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol,* 21**,** 177-84.

IACOPETTA, B., GRIEU, F. & AMANUEL, B. 2010. Microsatellite instability in colorectal cancer. *Asia Pac J Clin Oncol,* 6**,** 260-9.

IBRAHIM, S. 2014. *Proteins, which are upregulated at early time points following Apc deletion, are involved in intestinal tumourigenesis and represent potential colorectal cancer biomarkers.* PhD, University of Liverpool.

ILYAS, M., GRABSCH, H., ELLIS, I. O., WOMACK, C., BROWN, R., BERNEY, D., FENNELL, D., SALTO-TELLEZ, M., JENKINS, M., LANDBERG, G., BYERS, R., TREANOR, D., HARRISON, D., GREEN, A. R., BALL, G. & HAMILTON, P. 2013. Guidelines and considerations for conducting experiments using tissue microarrays. *Histopathology,* 62**,** 827-39.

ILYAS, M., TOMLINSON, I. P., ROWAN, A., PIGNATELLI, M. & BODMER, W. F. 1997. Beta-catenin mutations in cell lines established from human colorectal cancers. *Proc Natl Acad Sci U S A,* 94**,** 10330-4.

IMPERIALE, T. F., RANSOHOFF, D. F. & ITZKOWITZ, S. H. 2014. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med,* 371**,** 187-8.

IMPERIALE, T. F., RANSOHOFF, D. F., ITZKOWITZ, S. H., TURNBULL, B. A., ROSS, M. E. & GROUP, C. C. S. 2004. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med,* 351**,** 2704-14.

INCA 2011. Estimativa 2012 - Incidência de câncer no Brasil. 2011 ed.: Instituto Nacional do Cancer.

JACOBS, J. P., MAVROUDIS, C., JACOBS, M. L., MARUSZEWSKI, B., TCHERVENKOV, C. I., LACOUR-GAYET, F. G., CLARKE, D. R., YEH, T., JR., WALTERS, H. L., 3RD, KUROSAWA, H., STELLIN, G., EBELS, T. & ELLIOTT, M. J. 2006. What is operative mortality? Defining death in a surgical registry database: a report of the STS Congenital Database Taskforce and the Joint EACTS-STS Congenital Database Committee. *Ann Thorac Surg,* 81**,** 1937-41.

JARAJ, S. J., CAMPARO, P., BOYLE, H., GERMAIN, F., NILSSON, B., PETERSSON, F. & EGEVAD, L. 2009. Intra- and interobserver reproducibility of interpretation of immunohistochemical stains of prostate cancer. *Virchows Arch,* 455**,** 375-81.

JASS, J. R. 2007. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology,* 50**,** 113-30.

JEMAL, A., BRAY, F., CENTER, M. M., FERLAY, J., WARD, E. & FORMAN, D. 2011. Global cancer statistics. *CA Cancer J Clin,* 61**,** 69-90.

JIA, L., REN, J. M., WANG, Y. Y., ZHENG, Y., ZHANG, H., ZHANG, Q., KONG, B. H. & ZHENG, W. X. 2014. Inhibitory role of prohibitin in human ovarian epithelial cancer. *Int J Clin Exp Pathol,* 7**,** 2247-55.

JIANG, P., XIANG, Y., WANG, Y. J., LI, S. M., WANG, Y., HUA, H. R., YU, G. Y., ZHANG, Y., LEE, W. H. & ZHANG, Y. 2013. Differential expression and subcellular localization of Prohibitin 1 are related to tumorigenesis and progression of non-small cell lung cancer. *Int J Clin Exp Pathol,* 6**,** 2092-101.

JOHN, S. K., GEORGE, S., PRIMROSE, J. N. & FOZARD, J. B. 2011. Symptoms and signs in patients with colorectal cancer. *Colorectal Dis,* 13**,** 17-25.

JOHNSON, C. D., CHEN, M. H., TOLEDANO, A. Y., HEIKEN, J. P., DACHMAN, A., KUO, M. D., MENIAS, C. O., SIEWERT, B., CHEEMA, J. I., OBREGON, R. G., FIDLER, J. L., ZIMMERMAN, P., HORTON, K. M., COAKLEY, K., IYER, R. B., HARA, A. K., HALVORSEN, R. A., JR., CASOLA, G., YEE, J., HERMAN, B. A., BURGART, L. J. & LIMBURG, P. J. 2008. Accuracy of CT colonography for detection of large adenomas and cancers. *N Engl J Med,* 359**,** 1207-17.

JOHNSON, D. A., BARCLAY, R. L., MERGENER, K., WEISS, G., KÖNIG, T., BECK, J. & POTTER, N. T. 2014. Plasma Septin9 versus fecal immunochemical testing for colorectal cancer screening: a prospective multicenter study. *PLoS One,* 9**,** e98238.

JONES, D. P. 2008. Radical-free biology of oxidative stress. *Am J Physiol Cell Physiol,* 295**,** C849-68.

JUN, S. H., KIM, T. G. & BAN, C. 2006. DNA mismatch repair system. Classical and fresh roles. *FEBS J,* 273**,** 1609-19.

KAMANGAR, F., DORES, G. M. & ANDERSON, W. F. 2006. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol,* 24**,** 2137-50.

KANG, R., CHEN, R., ZHANG, Q., HOU, W., WU, S., CAO, L., HUANG, J., YU, Y., FAN, X. G., YAN, Z., SUN, X., WANG, H., WANG, Q., TSUNG, A., BILLIAR, T. R., ZEH, H. J., LOTZE, M. T. & TANG, D. 2014. HMGB1 in health and disease. *Mol Aspects Med*.

KANG, R., ZHANG, Q., ZEH, H. J., LOTZE, M. T. & TANG, D. 2013. HMGB1 in cancer: good, bad, or both? *Clin Cancer Res,* 19**,** 4046-57.

KANG, X., ZHANG, L., SUN, J., NI, Z., MA, Y., CHEN, X., SHENG, X. & CHEN, T. 2008. Prohibitin: a potential biomarker for tissue-based detection of gastric cancer. *J Gastroenterol,* 43**,** 618-25.

KAPTAIN, S., TAN, L. K. & CHEN, B. 2001. Her-2/neu and breast cancer. *Diagn Mol Pathol,* 10**,** 139-52.

KASASHIMA, K., SUMITANI, M., SATOH, M. & ENDO, H. 2008. Human prohibitin 1 maintains the organization and stability of the mitochondrial nucleoids. *Exp Cell Res,* 314**,** 988-96.

KATANODA, K., MATSUDA, T., MATSUDA, A., SHIBATA, A., NISHINO, Y., FUJITA, M., SODA, M., IOKA, A., SOBUE, T. & NISHIMOTO, H. 2013. An updated report of the trends in cancer incidence and mortality in Japan. *Jpn J Clin Oncol,* 43**,** 492-507.

KERR, S. E., THOMAS, C. B., THIBODEAU, S. N., FERBER, M. J. & HALLING, K. C. 2013. APC germline mutations in individuals being evaluated for familial adenomatous polyposis: a review of the Mayo Clinic experience with 1591 consecutive tests. *J Mol Diagn,* 15**,** 31-43.

KEWENTER, J., BREVINGE, H., ENGARÅS, B., HAGLIND, E. & AHRÉN, C. 1994. Results of screening, rescreening, and follow-up in a prospective randomized study for detection of colorectal cancer by fecal occult blood testing. Results for 68,308 subjects. *Scand J Gastroenterol,* 29**,** 468-73.

KHEIRELSEID, E. A., CHANG, K. H., NEWELL, J., KERIN, M. J. & MILLER, N. 2010. Identification of endogenous control genes for normalisation of real-time quantitative PCR data in colorectal cancer. *BMC Mol Biol,* 11**,** 12.

KIDD, M., MODLIN, I. M., MANE, S. M., CAMP, R. L., EICK, G. & LATICH, I. 2006. The role of genetic markers--NAP1L1, MAGE-D2, and MTA1--in defining small-intestinal carcinoid neoplasia. *Ann Surg Oncol,* 13**,** 253-62.

KIM, D. H., PICKHARDT, P. J., TAYLOR, A. J., LEUNG, W. K., WINTER, T. C., HINSHAW, J. L., GOPAL, D. V., REICHELDERFER, M., HSU, R. H. & PFAU, P. R. 2007a. CT colonography

versus colonoscopy for the detection of advanced neoplasia. *N Engl J Med,* 357**,** 1403-12.
KIM, S. Y., KIM, Y., HWANG, H. Y. & KIM, T. Y. 2007b. Altered expression of prohibitin in psoriatic lesions and its cellular implication. *Biochem Biophys Res Commun,* 360**,** 653-8.
KINZLER, K. W., NILBERT, M. C., SU, L. K., VOGELSTEIN, B., BRYAN, T. M., LEVY, D. B., SMITH, K. J., PREISINGER, A. C., HEDGE, P. & MCKECHNIE, D. 1991. Identification of FAP locus genes from chromosome 5q21. *Science,* 253**,** 661-5.
KINZLER, K. W. & VOGELSTEIN, B. 1997. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature,* 386**,** 761, 763.
KOBAYASHI, M., HONMA, T., MATSUDA, Y., SUZUKI, Y., NARISAWA, R., AJIOKA, Y. & ASAKURA, H. 2000. Nuclear translocation of beta-catenin in colorectal cancer. *Br J Cancer,* 82**,** 1689-93.
KOPETZ, S., TABERNERO, J., ROSENBERG, R., JIANG, Z. Q., MORENO, V., BACHLEITNER-HOFMANN, T., LANZA, G., STORK-SLOOTS, L., MARU, D., SIMON, I., CAPELLA, G. & SALAZAR, R. 2015. Genomic classifier ColoPrint predicts recurrence in stage II colorectal cancer patients more accurately than clinical factors. *Oncologist,* 20**,** 127-33.
KOSTI, I., JAIN, N., ARAN, D., BUTTE, A. J. & SIROTA, M. 2016. Cross-tissue Analysis of Gene and Protein Expression in Normal and Cancer Tissues. *Sci Rep,* 6**,** 24799.
KRONBORG, O., FENGER, C., OLSEN, J., JORGENSEN, O. D. & SONDERGAARD, O. 1996. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet,* 348**,** 1467-71.
KUHLE, J., BARRO, C., ANDREASSON, U., DERFUSS, T., LINDBERG, R., SANDELIUS, A., LIMAN, V., NORGREN, N., BLENNOW, K. & ZETTERBERG, H. 2016. Comparison of three analytical platforms for quantification of the neurofilament light chain in blood samples: ELISA, electrochemiluminescence immunoassay and Simoa. *Clin Chem Lab Med*.
KUIPERS, E. J. 2014. Colorectal cancer: screening-one small step for mankind, one giant leap for man. *Nat Rev Clin Oncol,* 11**,** 5-6.
KUIPERS, E. J., RÖSCH, T. & BRETTHAUER, M. 2013. Colorectal cancer screening--optimizing current strategies and new directions. *Nat Rev Clin Oncol,* 10**,** 130-42.
KUNTZ, K. M., LANSDORP-VOGELAAR, I., RUTTER, C. M., KNUDSEN, A. B., VAN BALLEGOOIJEN, M., SAVARINO, J. E., FEUER, E. J. & ZAUBER, A. G. 2011. A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. *Med Decis Making,* 31**,** 530-9.
KUROME, M., KATO, J., NAWA, T., FUJIMOTO, T., YAMAMOTO, H., SHIODE, J., WATO, M., KUWAKI, K., OKADA, H., SAKAGUCHI, K. & SHIRATORI, Y. 2008. Risk factors for high-grade dysplasia or carcinoma in colorectal adenoma cases treated with endoscopic polypectomy. *Eur J Gastroenterol Hepatol,* 20**,** 111-7.
KUROSUMI, M. 2003. Significance of immunohistochemical assessment of steroid hormone receptor status for breast cancer patients. *Breast Cancer,* 10**,** 97-104.
LA VECCHIA, C., BOSETTI, C., LUCCHINI, F., BERTUCCIO, P., NEGRI, E., BOYLE, P. & LEVI, F. 2010. Cancer mortality in Europe, 2000-2004, and an overview of trends since 1975. *Ann Oncol,* 21**,** 1323-60.
LAI, M. D. & XU, J. 2007. Ribosomal proteins and colorectal cancer. *Curr Genomics,* 8**,** 43-9.
LANGLEY, R. R. & FIDLER, I. J. 2011. The seed and soil hypothesis revisited--the role of tumor-stroma interactions in metastasis to different organs. *Int J Cancer,* 128**,** 2527-35.
LEARY, B. A., LAWRENCE-HENDERSON, R., MALLOZZI, C., FERNANDEZ OCANA, M., DURIGA, N., O'HARA, D. M., KAVOSI, M., QU, Q. & JOYCE, A. P. 2013. Bioanalytical platform

comparison using a generic human IgG PK assay format. *J Immunol Methods,* 397**,** 28-36.
LECLERC, D., DEJGAARD, K., MAZUR, A., DENG, L., WU, Q., NILSSON, T. & ROZEN, R. 2014. Quantitative proteomics reveals differentially expressed proteins in murine preneoplastic intestine in a model of intestinal tumorigenesis induced by low dietary folate and MTHFR deficiency. *Proteomics*. 2014, 14, 2558–2565.
LEE, H., SONG, M., SHIN, N., SHIN, C. H., MIN, B. S., KIM, H. S., YOO, J. S. & KIM, H. 2012. Diagnostic significance of serum HMGB1 in colorectal carcinomas. *PLoS One,* 7**,** e34318.
LEEDHAM, S. J. & WRIGHT, N. A. 2008. Expansion of a mutated clone: from stem cell to tumour. *J Clin Pathol,* 61**,** 164-71.
LEEMAN, M. F., MCKAY, J. A. & MURRAY, G. I. 2002. Matrix metalloproteinase 13 activity is associated with poor prognosis in colorectal cancer. *J Clin Pathol,* 55**,** 758-62.
LEMMON, M. A. & SCHLESSINGER, J. 2010. Cell signaling by receptor tyrosine kinases. *Cell,* 141**,** 1117-34.
LEQUIN, R. M. 2005. Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clin Chem,* 51**,** 2415-8.
LESLIE, A., CAREY, F. A., PRATT, N. R. & STEELE, R. J. 2002. The colorectal adenoma-carcinoma sequence. *Br J Surg,* 89**,** 845-60.
LEURAUD, K., JEZEWSKI-SERRA, D., VIGUIER, J. & SALINES, E. 2013. Colorectal cancer screening by guaiac faecal occult blood test in France: Evaluation of the programme two years after launching. *Cancer Epidemiol,* 37**,** 959-67.
LEVIN, B., LIEBERMAN, D. A., MCFARLAND, B., ANDREWS, K. S., BROOKS, D., BOND, J., DASH, C., GIARDIELLO, F. M., GLICK, S., JOHNSON, D., JOHNSON, C. D., LEVIN, T. R., PICKHARDT, P. J., REX, D. K., SMITH, R. A., THORSON, A. & WINAWER, S. J. 2008. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology,* 134**,** 1570-95.
LEVINE, A. J. & OREN, M. 2009. The first 30 years of p53: growing ever more complex. *Nat Rev Cancer,* 9**,** 749-58.
LEVY, M. S. F. 1974. O papel da migração internacional na evolução da população brasileira (1872 a 1972). *Revista de Saúde Pública,* 8**,** 49-90.
LI, D. & CHAN, D. W. 2014. Proteomic cancer biomarkers from discovery to approval: it's worth the effort. *Expert Rev Proteomics,* 11**,** 135-6.
LI, J., MERCER, E., GOU, X. & LU, Y. J. 2013a. Ethnical disparities of prostate cancer predisposition: genetic polymorphisms in androgen-related genes. *Am J Cancer Res,* 3**,** 127-51.
LI, L., GONG, H., YU, H., LIU, X., LIU, Q., YAN, G., ZHANG, Y., LU, H., ZOU, Y. & YANG, P. 2012. Knockdown of nucleosome assembly protein 1-like 1 promotes dimethyl sulfoxide-induced differentiation of P19CL6 cells into cardiomyocytes. *J Cell Biochem,* 113**,** 3788-96.
LI, L. F., WEI, Z. J., SUN, H. & JIANG, B. 2014a. Abnormal beta-catenin immunohistochemical expression as a prognostic factor in gastric cancer: a meta-analysis. *World J Gastroenterol,* 20**,** 12313-21.
LI, S., SUN, Y. & LI, L. 2014b. The expression of beta-catenin in different subtypes of breast cancer and its clinical significance. *Tumour Biol,* 35**,** 7693-8.
LI, X., YAO, X., WANG, Y., HU, F., WANG, F., JIANG, L., LIU, Y., WANG, D., SUN, G. & ZHAO, Y. 2013b. MLH1 promoter methylation frequency in colorectal cancer patients and related clinicopathological and molecular features. *PLoS One,* 8**,** e59064.
LIAO, Q., GUO, X., LI, X., XIONG, W., LI, X., YANG, J., CHEN, P., ZHANG, W., YU, H., TANG, H., DENG, M., LIANG, F., WU, M., LUO, Z., WANG, R., ZENG, X., ZENG, Z. & LI, G. 2013.

Prohibitin is an important biomarker for nasopharyngeal carcinoma progression and prognosis. *Eur J Cancer Prev,* 22**,** 68-76.
LIN, F. & CHEN, Z. 2014. Standardization of diagnostic immunohistochemistry: literature review and geisinger experience. *Arch Pathol Lab Med,* 138**,** 1564-77.
LIN, S., COUTINHO-MANSFIELD, G., WANG, D., PANDIT, S. & FU, X. D. 2008. The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol,* 15**,** 819-26.
LINE, A., SLUCKA, Z., STENGREVICS, A., SILINA, K., LI, G. & REES, R. C. 2002. Characterisation of tumour-associated antigens in colon cancer. *Cancer Immunol Immunother,* 51**,** 574-82.
LIU, Z., DING, Y., YE, N., WILD, C., CHEN, H. & ZHOU, J. 2016. Direct Activation of Bax Protein for Cancer Therapy. *Med Res Rev,* 36**,** 313-41.
LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.
LOUGHREY, M. B., QUIRKE, P. & SHEPHERD, N. A. 2014. *Standards and datasets for reporting cancers - Dataset for colorectal cancer histopathology reports* [Online]. London, UK: Royal College of Pathologists. Available: http://www.rcpath.org/Resources/RCPath/Migrated%20Resources/Documents/G/G049_ColorectalDataset_July14.pdf [Last accessed 05/08/15].
LOWE, S. W., CEPERO, E. & EVAN, G. 2004. Intrinsic tumour suppression. *Nature,* 432**,** 307-15.
LU, X. Y., LU, Y., ZHAO, Y. J., JAEWEON, K., KANG, J., XIAO-NAN, L., GE, G., MEYER, R., PERLAKY, L., HICKS, J., CHINTAGUMPALA, M., CAI, W. W., LADANYI, M., GORLICK, R., LAU, C. C., PATI, D., SHELDON, M. & RAO, P. H. 2008. Cell cycle regulator gene CDC5L, a potential target for 6p12-p21 amplicon in osteosarcoma. *Mol Cancer Res,* 6**,** 937-46.
LYNCH, H. T., SHAW, M. W., MAGNUSON, C. W., LARSEN, A. L. & KRUSH, A. J. 1966. Hereditary factors in cancer. Study of two large midwestern kindreds. *Arch Intern Med,* 117**,** 206-12.
LYNCH, H. T., WATSON, P., SHAW, T. G., LYNCH, J. F., HARTY, A. E., FRANKLIN, B. A., KAPLER, C. R., TINLEY, S. T. & LIU, B. 1999. Clinical impact of molecular genetic diagnosis, genetic counseling, and management of hereditary cancer. Part I: Studies of cancer in families. *Cancer,* 86**,** 2449-56.
MACARAEG, C., ORTIZ, J., CALAMBA, D. & MA, M. 2015. Serum sample stability in ligand-binding assays: challenges in assessments of long-term, bench-top and multiple freeze-thaw. *Bioanalysis,* 7**,** 1361-70.
MAJEK, O., GONDOS, A., JANSEN, L., EMRICH, K., HOLLECZEK, B., KATALINIC, A., NENNECKE, A., EBERLE, A. & BRENNER, H. 2012. Survival from colorectal cancer in Germany in the early 21st century. *Br J Cancer,* 106**,** 1875-80.
MAJEK, O., GONDOS, A., JANSEN, L., EMRICH, K., HOLLECZEK, B., KATALINIC, A., NENNECKE, A., EBERLE, A. & BRENNER, H. 2013. Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in Germany. *PLoS One,* 8**,** e68077.
MALMSTROM, P. U., WESTER, K., VASKO, J. & BUSCH, C. 1992. Expression of proliferative cell nuclear antigen (PCNA) in urinary bladder carcinoma. Evaluation of antigen retrieval methods. *Apmis,* 100**,** 988-92.
MALUMBRES, M. & BARBACID, M. 2003. RAS oncogenes: the first 30 years. *Nat Rev Cancer,* 3**,** 459-65.
MANDEL, J. S., BOND, J. H., CHURCH, T. R., SNOVER, D. C., BRADLEY, G. M., SCHUMAN, L. M. & EDERER, F. 1993. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N Engl J Med,* 328**,** 1365-71.
MANLEY, J. L. & TACKE, R. 1996. SR proteins and splicing control. *Genes Dev,* 10**,** 1569-79.

MARCHESE, R. D., PUCHALSKI, D., MILLER, P., ANTONELLO, J., HAMMOND, O., GREEN, T., RUBINSTEIN, L. J., CAULFIELD, M. J. & SIKKEMA, D. 2009. Optimization and validation of a multiplex, electrochemiluminescence-based detection assay for the quantitation of immunoglobulin G serotype-specific antipneumococcal antibodies in human serum. *Clin Vaccine Immunol,* 16**,** 387-96.

MARTIN, J. W., CHILTON-MACNEILL, S., KOTI, M., VAN WIJNEN, A. J., SQUIRE, J. A. & ZIELENSKA, M. 2014. Digital expression profiling identifies RUNX2, CDC5L, MDM2, RECQL4, and CDK4 as potential predictive biomarkers for neo-adjuvant chemotherapy response in paediatric osteosarcoma. *PLoS One,* 9**,** e95843.

MARX, A. H., THARUN, L., MUTH, J., DANCAU, A. M., SIMON, R., YEKEBAS, E., KAIFI, J. T., MIRLACHER, M., BRUMMENDORF, T. H., BOKEMEYER, C., IZBICKI, J. R. & SAUTER, G. 2009. HER-2 amplification is highly homogenous in gastric cancer. *Hum Pathol,* 40**,** 769-77.

MATHERS, C. D., FAT, D. M. & BOERMA, J. T. 2008. *The global burden of disease: 2004 update*, World Health Organization.

MATHERS, C. D., FAT, D. M., INOUE, M., RAO, C. & LOPEZ, A. D. 2005. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull World Health Organ,* 83**,** 171-7.

MCCARTHY, D. J. & SMYTH, G. K. 2009. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics,* 25**,** 765-71.

MCCARTY, K. S., SZABO, E., FLOWERS, J. L., COX, E. B., LEIGHT, G. S., MILLER, L., KONRATH, J., SOPER, J. T., BUDWIT, D. A. & CREASMAN, W. T. 1986. Use of a monoclonal anti-estrogen receptor antibody in the immunohistochemical evaluation of human tumors. *Cancer Res,* 46**,** 4244s-4248s.

MCLEAN, M. H., MURRAY, G. I., STEWART, K. N., NORRIE, G., MAYER, C., HOLD, G. L., THOMSON, J., FYFE, N., HOPE, M., MOWAT, N. A., DREW, J. E. & EL-OMAR, E. M. 2011. The inflammatory microenvironment in colorectal neoplasia. *PLoS One,* 6**,** e15366.

MERDZHANOVA, G., EDMOND, V., DE SERANNO, S., VAN DEN BROECK, A., CORCOS, L., BRAMBILLA, C., BRAMBILLA, E., GAZZERI, S. & EYMIN, B. 2008. E2F1 controls alternative splicing pattern of genes involved in apoptosis through upregulation of the splicing factor SC35. *Cell Death Differ,* 15**,** 1815-23.

MERDZHANOVA, G., GOUT, S., KERAMIDAS, M., EDMOND, V., COLL, J. L., BRAMBILLA, C., BRAMBILLA, E., GAZZERI, S. & EYMIN, B. 2010. The transcription factor E2F1 and the SR protein SC35 control the ratio of pro-angiogenic versus antiangiogenic isoforms of vascular endothelial growth factor-A to inhibit neovascularization in vivo. *Oncogene,* 29**,** 5392-403.

MIHAJLOVIĆ, J., PECHLIVANOGLOU, P., MILADINOV-MIKOV, M., ZIVKOVIĆ, S. & POSTMA, M. J. 2013. Cancer incidence and mortality in Serbia 1999-2009. *BMC Cancer,* 13**,** 18.

MILANES-YEARSLEY, M., HAMMOND, M. E., PAJAK, T. F., COOPER, J. S., CHANG, C., GRIFFIN, T., NELSON, D., LARAMORE, G. & PILEPICH, M. 2002. Tissue micro-array: a cost and time-effective method for correlative studies by regional and national cancer study groups. *Mod Pathol,* 15**,** 1366-73.

MILLER, S. & STEELE, S. 2012. Novel molecular screening approaches in colorectal cancer. *J Surg Oncol,* 105**,** 459-67.

MIYATA, Y., KUMAGAI, K., NAGAOKA, T., KITAURA, K., KANEDA, G., KANAZAWA, H., SUZUKI, S., HAMADA, Y. & SUZUKI, R. 2015. Clinicopathological significance and prognostic value of Wilms' tumor gene expression in colorectal cancer. *Cancer Biomark,* 15**,** 789-97.

MOCELLIN, S., ROSSI, C. R., PILATI, P., NITTI, D. & MARINCOLA, F. M. 2003. Quantitative real-time PCR: a powerful ally in cancer research. *Trends Mol Med,* 9**,** 189-95.

MOLLER, H., SANDIN, F., ROBINSON, D., BRAY, F., KLINT, S., LINKLATER, K. M., LAMBERT, P. C., PAHLMAN, L., HOLMBERG, L. & MORRIS, E. 2012. Colorectal cancer survival in socioeconomic groups in England: variation is mainly in the short term after diagnosis. *Eur J Cancer,* 48**,** 46-53.

MORIKAWA, T., KUCHIBA, A., YAMAUCHI, M., MEYERHARDT, J. A., SHIMA, K., NOSHO, K., CHAN, A. T., GIOVANNUCCI, E., FUCHS, C. S. & OGINO, S. 2011. Association of CTNNB1 (beta-catenin) alterations, body mass index, and physical activity with survival in patients with colorectal cancer. *JAMA,* 305**,** 1685-94.

MORIN, P. J., SPARKS, A. B., KORINEK, V., BARKER, N., CLEVERS, H., VOGELSTEIN, B. & KINZLER, K. W. 1997. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science,* 275**,** 1787-90.

MORRIS, E. J., SANDIN, F., LAMBERT, P. C., BRAY, F., KLINT, A., LINKLATER, K., ROBINSON, D., PAHLMAN, L., HOLMBERG, L. & MOLLER, H. 2011. A population-based comparison of the survival of patients with colorectal cancer in England, Norway and Sweden between 1996 and 2004. *Gut,* 60**,** 1087-93.

MORRIS, M., PLATELL, C., DE BOER, B., MCCAUL, K. & IACOPETTA, B. 2006. Population-based study of prognostic factors in stage II colonic cancer. *Br J Surg,* 93**,** 866-71.

MOSER, A. R., PITOT, H. C. & DOVE, W. F. 1990. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science,* 247**,** 322-4.

MOURADOV, D., SLOGGETT, C., JORISSEN, R. N., LOVE, C. G., LI, S., BURGESS, A. W., ARANGO, D., STRAUSBERG, R. L., BUCHANAN, D., WORMALD, S., O'CONNOR, L., WILDING, J. L., BICKNELL, D., TOMLINSON, I. P., BODMER, W. F., MARIADASON, J. M. & SIEBER, O. M. 2014. Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res,* 74**,** 3238-47.

MU, K., LI, L., YANG, Q., YUN, H., KHARAZIHA, P., YE, D. W., AUER, G., LAGERCRANTZ, S. B. & ZETTERBERG, A. 2015. A standardized method for quantifying proliferation by Ki-67 and cyclin A immunohistochemistry in breast cancer. *Ann Diagn Pathol,* 19**,** 243-8.

MU, R., WANG, Y. B., WU, M., YANG, Y., SONG, W., LI, T., ZHANG, W. N., TAN, B., LI, A. L., WANG, N., XIA, Q., GONG, W. L., WANG, C. G., ZHOU, T., GUO, N., SANG, Z. H. & LI, H. Y. 2014. Depletion of pre-mRNA splicing factor Cdc5L inhibits mitotic progression and triggers mitotic catastrophe. *Cell Death Dis,* 5**,** e1151.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G. & ERLICH, H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol,* 51 Pt 1**,** 263-73.

MUMM, J. B. & OFT, M. 2008. Cytokine-based transformation of immune surveillance into tumor-promoting inflammation. *Oncogene,* 27**,** 5913-9.

MUTO, T., BUSSEY, H. J. & MORSON, B. C. 1975. The evolution of cancer of the colon and rectum. *Cancer,* 36**,** 2251-70.

MYANT, K. & SANSOM, O. J. 2011. Wnt/Myc interactions in intestinal cancer: partners in crime. *Exp Cell Res,* 317**,** 2725-31.

NAGATA, T., TAKAHASHI, Y., ISHII, Y., ASAI, S., NISHIDA, Y., MURATA, A., KOSHINAGA, T., FUKUZAWA, M., HAMAZAKI, M., ASAMI, K., ITO, E., IKEDA, H., TAKAMATSU, H., KOIKE, K., KIKUTA, A., KUROIWA, M., WATANABE, A., KOSAKA, Y., FUJITA, H., MIYAKE, M. & MUGISHIMA, H. 2003. Transcriptional profiling in hepatoblastomas using high-density oligonucleotide DNA array. *Cancer Genet Cytogenet,* 145**,** 152-60.

NAJM, M. Z., ZAIDI, S., SIDDIQUI, W. A. & HUSAIN, S. A. 2013. Immunohistochemical expression and mutation study of Prohibitin gene in Indian female breast cancer cases. *Med Oncol,* 30**,** 614.

NAZEMALHOSSEINI MOJARAD, E., KUPPEN, P. J., AGHDAEI, H. A. & ZALI, M. R. 2013. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench,* 6**,** 120-8.

NEGRINI, S., GORGOULIS, V. G. & HALAZONETIS, T. D. 2010. Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol,* 11**,** 220-8.

NG, E. K., CHONG, W. W., JIN, H., LAM, E. K., SHIN, V. Y., YU, J., POON, T. C., NG, S. S. & SUNG, J. J. 2009. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut,* 58**,** 1375-81.

NGUYEN, M. T. & WEINBERG, D. S. 2016. Biomarkers in Colorectal Cancer Screening. *J Natl Compr Canc Netw,* 14**,** 1033-40.

NIJTMANS, L. G., DE JONG, L., ARTAL SANZ, M., COATES, P. J., BERDEN, J. A., BACK, J. W., MUIJSERS, A. O., VAN DER SPEK, H. & GRIVELL, L. A. 2000. Prohibitins act as a membrane-bound chaperone for the stabilization of mitochondrial proteins. *EMBO J,* 19**,** 2444-51.

NISHISHO, I., NAKAMURA, Y., MIYOSHI, Y., MIKI, Y., ANDO, H., HORII, A., KOYAMA, K., UTSUNOMIYA, J., BABA, S. & HEDGE, P. 1991. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science,* 253**,** 665-9.

O'NEAL, W. K., ANDERSON, W., BASTA, P. V., CARRETTA, E. E., DOERSCHUK, C. M., BARR, R. G., BLEECKER, E. R., CHRISTENSON, S. A., CURTIS, J. L., HAN, M. K., HANSEL, N. N., KANNER, R. E., KLEERUP, E. C., MARTINEZ, F. J., MILLER, B. E., PETERS, S. P., RENNARD, S. I., SCHOLAND, M. B., TAL-SINGER, R., WOODRUFF, P. G., COUPER, D. J., DAVIS, S. M. & INVESTIGATORS, S. 2014. Comparison of serum, EDTA plasma and P100 plasma for luminex-based biomarker multiplex assays in patients with chronic obstructive pulmonary disease in the SPIROMICS study. *J Transl Med,* 12**,** 9.

OFFICE FOR NATIONAL STATISTICS. 2012. *ONS - Cancer Incidence and Mortality in the United Kingdom, 2008-10* [Online]. Office for National Statistics. Available: http://www.ons.gov.uk/ons/rel/cancer-unit/cancer-incidence-and-mortality/2008-2010/stb-cancer-incidence-and-mortality-in-the-united-kindom--2008-2010.html [Accessed 18 Oct 2013].

OFT, M. 2014. IL-10: master switch from tumor-promoting inflammation to antitumor immunity. *Cancer Immunol Res,* 2**,** 194-9.

OHMORI, H., LUO, Y. & KUNIYASU, H. 2011. Non-histone nuclear factor HMGB1 as a therapeutic target in colorectal cancer. *Expert Opin Ther Targets,* 15**,** 183-93.

OIEN, K. A. & DENNIS, J. L. 2012. Diagnostic work-up of carcinoma of unknown primary: from immunohistochemistry to molecular profiling. *Ann Oncol,* 23 Suppl 10**,** x271-7.

OKUWAKI, M., KATO, K. & NAGATA, K. 2010. Functional characterization of human nucleosome assembly protein 1-like proteins as histone chaperones. *Genes Cells,* 15**,** 13-27.

OLIVEIRA, T. C., SILVA, A. A., SANTOS CDE, J., SILVA, J. S. & CONCEICAO, S. I. 2010. Physical activity and sedentary lifestyle among children from private and public schools in Northern Brazil. *Rev Saude Publica,* 44**,** 996-1004.

OORT, F. A., TERHAAR SIVE DROSTE, J. S., VAN DER HULST, R. W., VAN HEUKELEM, H. A., LOFFELD, R. J., WESDORP, I. C., VAN WANROOIJ, R. L., DE BAAIJ, L., MUTSAERS, E. R., VAN DER REIJT, S., COUPE, V. M., BERKHOF, J., BOUMAN, A. A., MEIJER, G. A. & MULDER, C. J. 2010. Colonoscopy-controlled intra-individual comparisons to screen relevant neoplasia: faecal immunochemical test vs. guaiac-based faecal occult blood test. *Aliment Pharmacol Ther,* 31**,** 432-9.

PAGÈS, F., GALON, J., DIEU-NOSJEAN, M. C., TARTOUR, E., SAUTÈS-FRIDMAN, C. & FRIDMAN, W. H. 2010. Immune infiltration in human tumors: a prognostic factor that should not be ignored. *Oncogene,* 29**,** 1093-102.

PAIK, S., SHAK, S., TANG, G., KIM, C., BAKER, J., CRONIN, M., BAEHNER, F. L., WALKER, M. G., WATSON, D., PARK, T., HILLER, W., FISHER, E. R., WICKERHAM, D. L., BRYANT, J. & WOLMARK, N. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med,* 351**,** 2817-26.

PAN, Q., SHAI, O., LEE, L. J., FREY, B. J. & BLENCOWE, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet,* 40**,** 1413-5.

PARK, S. K., SONG, C. S., YANG, H. J., JUNG, Y. S., CHOI, K. Y., KOO, D. H., KIM, K. E., JEONG, K. U., KIM, H. O., KIM, H., CHUN, H. K. & PARK, D. I. 2016. Field Cancerization in Sporadic Colon Cancer. *Gut Liver*. 10(5):773-780.

PARK, Y. J. & LUGER, K. 2006. Structure and function of nucleosome assembly proteins. *Biochem Cell Biol,* 84**,** 549-58.

PARSONS, M. & GRABSCH, H. 2009. How to make tissue microarrays. *Diagnostic Histopathology,* 15**,** 142-150.

PECOT, C. V., CALIN, G. A., COLEMAN, R. L., LOPEZ-BERESTEIN, G. & SOOD, A. K. 2011. RNA interference in the clinic: challenges and future directions. *Nat Rev Cancer,* 11**,** 59-67.

PEEPLES, C., SHELLNUT, J., WASVARY, H., RIGGS, T. & SACKSNER, J. 2010. Predictive factors affecting survival in stage II colorectal cancer: is lymph node harvesting relevant? *Dis Colon Rectum,* 53**,** 1517-23.

PEKAR-ZLOTIN, M., HIRSCH, F. R., SOUSSAN-GUTMAN, L., ILOUZE, M., DVIR, A., BOYLE, T., WYNES, M., MILLER, V. A., LIPSON, D., PALMER, G. A., ALI, S. M., DEKEL, S., BRENNER, R., BUNN, P. A., JR. & PELED, N. 2015. Fluorescence in situ hybridization, immunohistochemistry, and next-generation sequencing for detection of EML4-ALK rearrangement in lung cancer. *Oncologist,* 20**,** 316-22.

PENG, R. Q., WU, X. J., DING, Y., LI, C. Y., YU, X. J., ZHANG, X., PAN, Z. Z., WAN, D. S., ZHENG, L. M., ZENG, Y. X. & ZHANG, X. S. 2010. Co-expression of nuclear and cytoplasmic HMGB1 is inversely associated with infiltration of CD45RO+ T cells and prognosis in patients with stage IIIB colon cancer. *BMC Cancer,* 10**,** 496.

PENGJUN, Z., XINYU, W., FENG, G., XINXIN, D., YULAN, L., JUAN, L., XINGWANG, J., ZHENNAN, D. & YAPING, T. 2013. Multiplexed cytokine profiling of serum for detection of colorectal cancer. *Future Oncol,* 9**,** 1017-27.

PEREZ, R. O., PROSCURSHIM, I., SÃO JULIÃO, G. P., PICOLO, M., GAMA-RODRIGUES, J. & HABR-GAMA, A. 2008. Instalação e resultados preliminares de programa de rastreamento populacional de câncer colorretal em município brasileiro. *ABCD. Arquivos Brasileiros de Cirurgia Digestiva (São Paulo),* 21**,** 12-15.

PETERSEN, V. C., BAXTER, K. J., LOVE, S. B. & SHEPHERD, N. A. 2002. Identification of objective pathological prognostic determinants and models of prognosis in Dukes' B colon cancer. *Gut,* 51**,** 65-9.

PINTO, L. F. & SORANZ, D. R. 2004. Planos privados de assistência à saúde: cobertura populacional no Brasil. *Cien Saude Colet,* 9**,** 85-98.

PLANUTIS, K., PLANUTIENE, M. & HOLCOMBE, R. F. 2014. A novel signaling pathway regulates colon cancer angiogenesis through Norrin. *Sci Rep,* 4**,** 5630.

PLAWSKI, A., BANASIEWICZ, T., BORUN, P., KUBASZEWSKI, L., KROKOWICZ, P., SKRZYPCZAK-ZIELINSKA, M. & LUBINSKI, J. 2013. Familial adenomatous polyposis of the colon. *Hered Cancer Clin Pract,* 11**,** 15.

POSTELNEK, J., NEELY, R. J., ROBBINS, M. D., GLEASON, C. R., PETERSON, J. E. & PICCOLI, S. P. 2016. Development and Validation of Electrochemiluminescence Assays to Measure Free and Total sSLAMF7 in Human Serum in the Absence and Presence of Elotuzumab. *Aaps j,* 18**,** 989-99.

POX, C. P., ALTENHOFEN, L., BRENNER, H., THEILMEIER, A., VON STILLFRIED, D. & SCHMIEGEL, W. 2012. Efficacy of a nationwide screening colonoscopy program for colorectal cancer. *Gastroenterology,* 142**,** 1460-7 e2.

PRAGER, G. W. & POETTLER, M. 2012. Angiogenesis in cancer. Basic mechanisms and therapeutic advances. *Hamostaseologie,* 32**,** 105-14.

PRASAD, K., TIWARI, A., ILANTHODI, S., PRABHU, G. & PAI, M. 2011. Automation of immunohistochemical evaluation in breast cancer using image analysis. *World J Clin Oncol,* 2**,** 187-94.

PRENEN, H., VECCHIONE, L. & VAN CUTSEM, E. 2013. Role of targeted agents in metastatic colorectal cancer. *Target Oncol,* 8**,** 83-96.

PRICE, C. P. & NEWMAN, D. J. 1997. *Principles and practice of immunoassay*, London : Macmillan Reference Ltd., 1997. 2nd ed.

PRITZKER, K. P. 2015. Predictive and prognostic cancer biomarkers revisited. *Expert Rev Mol Diagn,* 15**,** 971-4.

PUGNIERE, P., BANZET, S., CHAILLOU, T., MOURET, C. & PEINNEQUIN, A. 2011. Pitfalls of reverse transcription quantitative polymerase chain reaction standardization: Volume-related inhibitors of reverse transcription. *Anal Biochem,* 415**,** 151-7.

QIAN, W., IQBAL, K., GRUNDKE-IQBAL, I., GONG, C. X. & LIU, F. 2011. Splicing factor SC35 promotes tau expression through stabilization of its mRNA. *FEBS Lett,* 585**,** 875-80.

QIN, Y. J., ZHOU, X. Y., CAI, S. J., YAN, G., ZHANG, T. M. & SHI, D. R. 2006. [Real-time PCR analysis of beta-catenin mRNA in sporadic colorectal cancers]. *Zhonghua Bing Li Xue Za Zhi,* 35**,** 535-9.

QUIRKE, P. & MORRIS, E. 2007. Reporting colorectal cancer. *Histopathology,* 50**,** 103-12.

QUIRKE, P., RISIO, M., LAMBERT, R., VON KARSA, L. & VIETH, M. 2012. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Quality assurance in pathology in colorectal cancer screening and diagnosis. *Endoscopy,* 44 Suppl 3**,** Se116-30.

RANA, T. M. 2007. Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol,* 8**,** 23-36.

RAUDENSKA, M., SZTALMACHOVA, M., GUMULEC, J., FOJTU, M., POLANSKA, H., BALVAN, J., FEITH, M., BINKOVA, H., HORAKOVA, Z., KOSTRICA, R., KIZEK, R. & MASARIK, M. 2015a. Prognostic significance of the tumour-adjacent tissue in head and neck cancers. *Tumour Biol***,** 9929-39.

RAUDENSKA, M., SZTALMACHOVA, M., GUMULEC, J., FOJTU, M., POLANSKA, H., BALVAN, J., FEITH, M., BINKOVA, H., HORAKOVA, Z., KOSTRICA, R., KIZEK, R. & MASARIK, M. 2015b. Prognostic significance of the tumour-adjacent tissue in head and neck cancers. *Tumour Biol*. (2015) 36:9929–9939

REES, C. J. & BEVAN, R. 2013. The National Health Service Bowel Cancer Screening Program: the early years. *Expert Rev Gastroenterol Hepatol,* 7**,** 421-37.

REX, D. K., LEHMAN, G. A., ULBRIGHT, T. M., SMITH, J. J., POUND, D. C., HAWES, R. H., HELPER, D. J., WIERSEMA, M. J., LANGEFELD, C. D. & LI, W. 1993. Colonic neoplasia in asymptomatic persons with negative fecal occult blood tests: influence of age, gender, and family history. *Am J Gastroenterol,* 88**,** 825-31.

RIFAI, N., GILLETTE, M. A. & CARR, S. A. 2006. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol,* 24**,** 971-83.

RISIO, M. 2010. The natural history of adenomas. *Best Pract Res Clin Gastroenterol,* 24**,** 271-80.

RIZZARDI, A. E., JOHNSON, A. T., VOGEL, R. I., PAMBUCCIAN, S. E., HENRIKSEN, J., SKUBITZ, A. P., METZGER, G. J. & SCHMECHEL, S. C. 2012. Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol,* 7**,** 42.

ROCKEY, D. C., PAULSON, E., NIEDZWIECKI, D., DAVIS, W., BOSWORTH, H. B., SANDERS, L., YEE, J., HENDERSON, J., HATTEN, P., BURDICK, S., SANYAL, A., RUBIN, D. T., STERLING, M., AKERKAR, G., BHUTANI, M. S., BINMOELLER, K., GARVIE, J., BINI, E. J., MCQUAID, K., FOSTER, W. L., THOMPSON, W. M., DACHMAN, A. & HALVORSEN, R.

2005. Analysis of air contrast barium enema, computed tomographic colonography, and colonoscopy: prospective comparison. *Lancet,* 365**,** 305-11.
RODRIGUEZ, O. C., CHOUDHURY, S., KOLUKULA, V., VIETSCH, E. E., CATANIA, J., PREET, A., REYNOSO, K., BARGONETTI, J., WELLSTEIN, A., ALBANESE, C. & AVANTAGGIATI, M. L. 2012. Dietary downregulation of mutant p53 levels via glucose restriction: mechanisms and implications for tumor therapy. *Cell Cycle,* 11**,** 4436-46.
ROLLOT, F., CHAUVENET, M., ROCHE, L., HAMZA, S., LEPAGE, C., FAIVRE, J. & BOUVIER, A. M. 2013. Long-term net survival in patients with colorectal cancer in France: an informative contribution of recent methodology. *Dis Colon Rectum,* 56**,** 1118-24.
ROSELL, R., ICHINOSE, Y., TARON, M., SARRIES, C., QUERALT, C., MENDEZ, P., SANCHEZ, J. M., NISHIYAMA, K., MORAN, T., CIRAUQUI, B., MATE, J. L., BESSE, B., REGUART, N., PEREZ, M. & SANCHEZ, J. J. 2005. Mutations in the tyrosine kinase domain of the EGFR gene associated with gefitinib response in non-small-cell lung cancer. *Lung Cancer,* 50**,** 25-33.
RUBIN, S. M. & SAGE, J. 2013. Defining a new vision for the retinoblastoma gene: report from the 3rd International Rb Meeting. *Cell Div,* 8**,** 13.
RUIFROK, A. C. & JOHNSTON, D. A. 2001. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol,* 23**,** 291-9.
RUIFROK, A. C., KATZ, R. L. & JOHNSTON, D. A. 2003. Comparison of quantification of histochemical staining by hue-saturation-intensity (HSI) transformation and color-deconvolution. *Appl Immunohistochem Mol Morphol,* 11**,** 85-91.
SACH, T. H. & WHYNES, D. K. 2009. Men and women: beliefs about cancer and about screening. *BMC Public Health,* 9**,** 431.
SAFA, A. R. 2016. Resistance to Cell Death and Its Modulation in Cancer Stem Cells. *Crit Rev Oncog,* 21**,** 203-219.
SAGAERT, X. 2014. Prognostic biomarkers in colorectal cancer: where do we stand? *Virchows Arch,* 464**,** 379-91.
SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIS, K. B. & ERLICH, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239**,** 487-91.
SAMEER, A. S., NISSAR, S. & FATIMA, K. 2014. Mismatch repair pathway: molecules, functions, and role in colorectal carcinogenesis. *Eur J Cancer Prev,* 23**,** 246-57.
SANCHEZ, M. J., PAYER, T., DE ANGELIS, R., LARRANAGA, N., CAPOCACCIA, R. & MARTINEZ, C. 2010. Cancer incidence and mortality in Spain: estimates and projections for the period 1981-2012. *Ann Oncol,* 21 Suppl 3**,** iii30-36.
SANJOAQUIN, M. A., CHOODARI-OSKOOEI, B., DOLBEAR, C., PUTCHA, V., SEHGAL, A., KEY, T. J. & MOLLER, H. 2007. Colorectal cancer incidence, mortality and survival in South-east England between 1972 and 2001. *Eur J Cancer Prev,* 16**,** 10-6.
SANSOM, O. J., MENIEL, V. S., MUNCAN, V., PHESSE, T. J., WILKINS, J. A., REED, K. R., VASS, J. K., ATHINEOS, D., CLEVERS, H. & CLARKE, A. R. 2007. Myc deletion rescues Apc deficiency in the small intestine. *Nature,* 446**,** 676-9.
SANSOM, O. J., REED, K. R., HAYES, A. J., IRELAND, H., BRINKMANN, H., NEWTON, I. P., BATLLE, E., SIMON-ASSMANN, P., CLEVERS, H., NATHKE, I. S., CLARKE, A. R. & WINTON, D. J. 2004. Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev,* 18**,** 1385-90.
SANZ-PAMPLONA, R., BERENGUER, A., CORDERO, D., MOLLEVI, D. G., CROUS-BOU, M., SOLE, X., PARE-BRUNET, L., GUINO, E., SALAZAR, R., SANTOS, C., DE OCA, J., SANJUAN, X., RODRIGUEZ-MORANTA, F. & MORENO, V. 2014. Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. *Mol Cancer,* 13**,** 46.

SAUS, E., BRUNET-VEGA, A., IRAOLA-GUZMAN, S., PEGUEROLES, C., GABALDON, T. & PERICAY, C. 2016. Long Non-Coding RNAs As Potential Novel Prognostic Biomarkers in Colorectal Cancer. *Front Genet,* 7**,** 54.

SCHIMMACK, S., TAYLOR, A., LAWRENCE, B., ALAIMO, D., SCHMITZ-WINNENTHAL, H., BUCHLER, M. W., MODLIN, I. M. & KIDD, M. 2014. A mechanistic role for the chromatin modulator, NAP1L1, in pancreatic neuroendocrine neoplasm proliferation and metastases. *Epigenetics Chromatin,* 7**,** 15.

SCHMANDT, R. & MILLS, G. B. 1993. Genomic components of carcinogenesis. *Clin Chem,* 39**,** 2375-85.

SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc,* 3**,** 1101-8.

SCHNEIDER, C. A., RASBAND, W. S. & ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* 9**,** 671-5.

SCHOEN, R. E., PINSKY, P. F., WEISSFELD, J. L., YOKOCHI, L. A., CHURCH, T., LAIYEMO, A. O., BRESALIER, R., ANDRIOLE, G. L., BUYS, S. S., CRAWFORD, E. D., FOUAD, M. N., ISAACS, C., JOHNSON, C. C., REDING, D. J., O'BRIEN, B., CARRICK, D. M., WRIGHT, P., RILEY, T. L., PURDUE, M. P., IZMIRLIAN, G., KRAMER, B. S., MILLER, A. B., GOHAGAN, J. K., PROROK, P. C., BERG, C. D. & TEAM, P. P. 2012. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med,* 366**,** 2345-57.

SCHOENFELD, P., CASH, B., FLOOD, A., DOBHAN, R., EASTONE, J., COYLE, W., KIKENDALL, J. W., KIM, H. M., WEISS, D. G., EMORY, T., SCHATZKIN, A. & LIEBERMAN, D. 2005. Colonoscopic screening of average-risk women for colorectal neoplasia. *N Engl J Med,* 352**,** 2061-8.

SCHOLEFIELD, J. H., MOSS, S. M., MANGHAM, C. M., WHYNES, D. K. & HARDCASTLE, J. D. 2012. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. *Gut,* 61**,** 1036-40.

SCHREIBER, R. D., OLD, L. J. & SMYTH, M. J. 2011. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science,* 331**,** 1565-70.

SCOTT, A., AMBANNAVAR, R., JEONG, J., LIU, M. L. & CRONIN, M. T. 2011. RT-PCR-based gene expression profiling for cancer biomarker discovery from fixed, paraffin-embedded tissues. *Methods Mol Biol,* 724**,** 239-57.

SEGNAN, N., ARMAROLI, P., BONELLI, L., RISIO, M., SCIALLERO, S., ZAPPA, M., ANDREONI, B., ARRIGONI, A., BISANTI, L., CASELLA, C., CROSTA, C., FALCINI, F., FERRERO, F., GIACOMIN, A., GIULIANI, O., SANTARELLI, A., VISIOLI, C. B., ZANETTI, R., ATKIN, W. S. & SENORE, C. 2011. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial--SCORE. *J Natl Cancer Inst,* 103**,** 1310-22.

SELLAMI, M., GAMOUDI, M., KRICHEN, K., KHARRAT, A., BEN ROMDHANE, K. & MAALEJ, M. 1991. [Incidence of amplification of the C-erb B2/Her-2/neu gene in human breast cancer]. *Arch Inst Pasteur Tunis,* 68**,** 33-41.

SEMENOV, M. V., HABAS, R., MACDONALD, B. T. & HE, X. 2007. SnapShot: Noncanonical Wnt Signaling Pathways. *Cell,* 131**,** 1378.

SHARMA, S., LIAO, W., ZHOU, X., WONG, D. T. & LICHTENSTEIN, A. 2011. Exon 11 skipping of E-cadherin RNA downregulates its expression in head and neck cancer cells. *Mol Cancer Ther,* 10**,** 1751-9.

SHAUKAT, A., MONGIN, S. J., GEISSER, M. S., LEDERLE, F. A., BOND, J. H., MANDEL, J. S. & CHURCH, T. R. 2013. Long-term mortality after screening for colorectal cancer. *N Engl J Med,* 369**,** 1106-14.

SHAW, P., BOVEY, R., TARDY, S., SAHLI, R., SORDAT, B. & COSTA, J. 1992. Induction of apoptosis by wild-type p53 in a human colon tumor-derived cell line. *Proc Natl Acad Sci U S A,* 89**,** 4495-9.

SHAY, G., LYNCH, C. C. & FINGLETON, B. 2015. Moving targets: Emerging roles for MMPs in cancer progression and metastasis. *Matrix Biol,* 44-46**,** 200-6.
SHAY, J. W. & WRIGHT, W. E. 2011. Role of telomeres and telomerase in cancer. *Semin Cancer Biol,* 21**,** 349-53.
SHERR, C. J. & MCCORMICK, F. 2002. The RB and p53 pathways in cancer. *Cancer Cell,* 2**,** 103-12.
SHERRILL, B., KAYE, J. A., SANDIN, R., CAPPELLERI, J. C. & CHEN, C. 2012. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther,* 5**,** 287-96.
SHIGEMATSU, H., LIN, L., TAKAHASHI, T., NOMURA, M., SUZUKI, M., WISTUBA, II, FONG, K. M., LEE, H., TOYOOKA, S., SHIMIZU, N., FUJISAWA, T., FENG, Z., ROTH, J. A., HERZ, J., MINNA, J. D. & GAZDAR, A. F. 2005. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst,* 97**,** 339-46.
SHU, J., DOLMAN, G. E., DUAN, J., QIU, G. & ILYAS, M. 2016. Statistical colour models: an automated digital image analysis method for quantification of histological biomarkers. *Biomed Eng Online,* 15**,** 46.
SHWEIKI, D., ITIN, A., SOFFER, D. & KESHET, E. 1992. Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. *Nature,* 359**,** 843-5.
SIEGEL, R. L., WARD, E. M. & JEMAL, A. 2012. Trends in colorectal cancer incidence rates in the United States by tumor location and stage, 1992-2008. *Cancer Epidemiol Biomarkers Prev,* 21**,** 411-6.
SILVA, G. A., GAMARRA, C. J., GIRIANELLI, V. R. & VALENTE, J. G. 2011. Cancer mortality trends in Brazilian state capitals and other municipalities between 1980 and 2006. *Rev Saude Publica,* 45**,** 1009-18.
SKEHAN, P., STORENG, R., SCUDIERO, D., MONKS, A., MCMAHON, J., VISTICA, D., WARREN, J. T., BOKESCH, H., KENNEY, S. & BOYD, M. R. 1990. New colorimetric cytotoxicity assay for anticancer-drug screening. *J Natl Cancer Inst,* 82**,** 1107-12.
SKRZYPSKI, M. 2008. Quantitative reverse transcriptase real-time polymerase chain reaction (qRT-PCR) in translational oncology: lung cancer perspective. *Lung Cancer,* 59**,** 147-54.
SLABY, O., SVOBODA, M., MICHALEK, J. & VYZULA, R. 2009. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. *Mol Cancer,* 8**,** 102.
SLAUGHTER, D. P., SOUTHWICK, H. W. & SMEJKAL, W. 1953. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer,* 6**,** 963-8.
SLOAN, J. H., SIEGEL, R. W., IVANOVA-COX, Y. T., WATSON, D. E., DEEG, M. A. & KONRAD, R. J. 2012. A novel high-sensitivity electrochemiluminescence (ECL) sandwich immunoassay for the specific quantitative measurement of plasma glucagon. *Clin Biochem,* 45**,** 1640-4.
SMITH, D. R. & GOH, H. S. 1996. Overexpression of the c-myc proto-oncogene in colorectal carcinoma is associated with a reduced mortality that is abrogated by point mutation of the p53 tumor suppressor gene. *Clin Cancer Res,* 2**,** 1049-53.
SODIR, N. M., SWIGART, L. B., KARNEZIS, A. N., HANAHAN, D., EVAN, G. I. & SOUCEK, L. 2011. Endogenous Myc maintains the tumor microenvironment. *Genes Dev,* 25**,** 907-16.
SOUCEK, L., WHITFIELD, J., MARTINS, C. P., FINCH, A. J., MURPHY, D. J., SODIR, N. M., KARNEZIS, A. N., SWIGART, L. B., NASI, S. & EVAN, G. I. 2008. Modelling Myc inhibition as a cancer therapy. *Nature,* 455**,** 679-83.
STEELE, R. J., MCCLEMENTS, P., WATLING, C., LIBBY, G., WELLER, D., BREWSTER, D. H., BLACK, R., CAREY, F. A. & FRASER, C. G. 2012. Interval cancers in a FOBT-based

colorectal cancer population screening programme: implications for stage, gender and tumour site. *Gut,* 61**,** 576-81.
STEINHAGEN, E., SHIA, J., MARKOWITZ, A. J., STADLER, Z. K., SALO-MULLEN, E. E., ZHENG, J., LEE-KONG, S. A., NASH, G. M., OFFIT, K. & GUILLEM, J. G. 2012. Systematic immunohistochemistry screening for Lynch syndrome in early age-of-onset colorectal cancer patients undergoing surgical resection. *J Am Coll Surg,* 214**,** 61-7.
STEPHEN, A. G., ESPOSITO, D., BAGNI, R. K. & MCCORMICK, F. 2014. Dragging ras back in the ring. *Cancer Cell,* 25**,** 272-81.
STOOP, E. M., DE HAAN, M. C., DE WIJKERSLOOTH, T. R., BOSSUYT, P. M., VAN BALLEGOOIJEN, M., NIO, C. Y., VAN DE VIJVER, M. J., BIERMANN, K., THOMEER, M., VAN LEERDAM, M. E., FOCKENS, P., STOKER, J., KUIPERS, E. J. & DEKKER, E. 2012. Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. *Lancet Oncol,* 13**,** 55-64.
STRUEWING, J. P., HARTGE, P., WACHOLDER, S., BAKER, S. M., BERLIN, M., MCADAMS, M., TIMMERMAN, M. M., BRODY, L. C. & TUCKER, M. A. 1997. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med,* 336**,** 1401-8.
STRYKER, S. J., WOLFF, B. G., CULP, C. E., LIBBE, S. D., ILSTRUP, D. M. & MACCARTY, R. L. 1987. Natural history of untreated colonic polyps. *Gastroenterology,* 93**,** 1009-13.
SU, L. K., KINZLER, K. W., VOGELSTEIN, B., PREISINGER, A. C., MOSER, A. R., LUONGO, C., GOULD, K. A. & DOVE, W. F. 1992. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science,* 256**,** 668-70.
SWANN, J. B. & SMYTH, M. J. 2007. Immune surveillance of tumors. *J Clin Invest,* 117**,** 1137-46.
SYSEL, A. M., VALLI, V. E., NAGLE, R. B. & BAUER, J. A. 2013. Immunohistochemical quantification of the vitamin B12 transport protein (TCII), cell surface receptor (TCII-R) and Ki-67 in human tumor xenografts. *Anticancer Res,* 33**,** 4203-12.
SØRBY, L. A., ANDERSEN, S. N., BUKHOLM, I. R. & JACOBSEN, M. B. 2010. Evaluation of suitable reference genes for normalization of real-time reverse transcription PCR analysis in colon cancer. *J Exp Clin Cancer Res,* 29**,** 144.
SÜREN, D., YILDIRIM, M., DEMIRPENÇE, Ö., KAYA, V., ALIKANOĞLU, A. S., BÜLBÜLLER, N., YILDIZ, M. & SEZER, C. 2014. The role of high mobility group box 1 (HMGB1) in colorectal cancer. *Med Sci Monit,* 20**,** 530-7.
TALMADGE, J. E. & FIDLER, I. J. 2010. AACR centennial series: the biology of cancer metastasis: historical perspective. *Cancer Res,* 70**,** 5649-69.
TAN, E., GOUVAS, N., NICHOLLS, R. J., ZIPRIN, P., XYNOS, E. & TEKKIS, P. P. 2009. Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. *Surg Oncol,* 18**,** 15-24.
TAYLOR, C. R. 2014. Immunohistochemistry in surgical pathology: principles and practice. *Methods Mol Biol,* 1180**,** 81-109.
TEMPLETON, N. S. 1992. The polymerase chain reaction. History, methods, and applications. *Diagn Mol Pathol,* 1**,** 58-72.
TETSU, O. & MCCORMICK, F. 1999. Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature,* 398**,** 422-6.
THIIS-EVENSEN, E., HOFF, G. S., SAUAR, J., LANGMARK, F., MAJAK, B. M. & VATN, M. H. 1999. Population-based surveillance by colonoscopy: effect on the incidence of colorectal cancer. Telemark Polyp Study I. *Scand J Gastroenterol,* 34**,** 414-20.
THOMAS, L. R. & TANSEY, W. P. 2011. Proteolytic control of the oncoprotein transcription factor Myc. *Adv Cancer Res,* 110**,** 77-106.

THUAUD, F., RIBEIRO, N., NEBIGIL, C. G. & DESAUBRY, L. 2013. Prohibitin ligands in cell death and survival: mode of action and therapeutic potential. *Chem Biol,* 20**,** 316-31.

TJALSMA, H. 2010. Identification of biomarkers for colorectal cancer through proteomics-based approaches. *Expert Rev Proteomics,* 7**,** 879-95.

TOFFART, A. C., TIMSIT, J. F., COURAUD, S., MERLE, P., MORO-SIBILOT, D., PEROL, M., MASTROIANNI, B., SOUQUET, P. J., GIRARD, N., JEANNIN, G., ROMAND, P., CHATELLAIN, P., VESIN, A., BRAMBILLA, C. & BRAMBILLA, E. 2014. Immunohistochemistry evaluation of biomarker expression in non-small cell lung cancer (Pharmacogenoscan study). *Lung Cancer,* 83**,** 182-8.

TOON, C. W., CHOU, A., CLARKSON, A., DESILVA, K., HOUANG, M., CHAN, J. C., SIOSON, L. L., JANKOVA, L. & GILL, A. J. 2014. Immunohistochemistry for myc predicts survival in colorectal cancer. *PLoS One,* 9**,** e87456.

TORRES, U. D. S., ALMEIDA, T. E. P. D. & NETINHO, J. G. 2010. Increasing hospital admission rates and economic burden for colorectal cancer in Brazil, 1996-2008. *Revista Panamericana de Salud Pública,* 28**,** 244-248.

TOTH, K., SIPOS, F., KALMAR, A., PATAI, A. V., WICHMANN, B., STOEHR, R., GOLCHER, H., SCHELLERER, V., TULASSAY, Z. & MOLNAR, B. 2012. Detection of methylated SEPT9 in plasma is a reliable screening method for both left- and right-sided colon cancers. *PLoS One,* 7**,** e46000.

TOYOTA, M., AHUJA, N., OHE-TOYOTA, M., HERMAN, J. G., BAYLIN, S. B. & ISSA, J. P. 1999. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A,* 96**,** 8681-6.

TRUANT, S. C., GOUYER, V. P., LETEURTRE, E. A., ZERIMECH, F., HUET, G. M. & PRUVOT, F. R. 2008. E-cadherin and beta-catenin mRNA levels throughout colon cancer progression. *J Surg Res,* 150**,** 212-8.

TSAI, W. S., HSIEH, P. S., YEH, C. Y., CHIANG, J. M., TANG, R., CHEN, J. S., CHANGCHIEN, C. R. & WANG, J. Y. 2011. Long-term survival benefits of adjuvant chemotherapy by decreasing incidence of tumor recurrence without delaying relapse in stage III colorectal cancer. *Int J Colorectal Dis,* 26**,** 1329-38.

TUOMINEN, V. J., RUOTOISTENMÄKI, S., VIITANEN, A., JUMPPANEN, M. & ISOLA, J. 2010. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res,* 12**,** R56.

UCHIDA, C. 2012. The retinoblastoma protein: functions beyond the G1-S regulator. *Curr Drug Targets,* 13**,** 1622-32.

UEDA, M., TAKAHASHI, Y., SHINDEN, Y., SAKIMURA, S., HIRATA, H., UCHI, R., TAKANO, Y., KURASHIGE, J., IGUCHI, T., EGUCHI, H., SUGIMACHI, K., YAMAMOTO, H., DOKI, Y., MORI, M. & MIMORI, K. 2014. Prognostic significance of high mobility group box 1 (HMGB1) expression in patients with colorectal cancer. *Anticancer Res,* 34**,** 5357-62.

UHLEN, M., FAGERBERG, L., HALLSTROM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, A., KAMPF, C., SJOSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S., ALM, T., EDQVIST, P. H., BERLING, H., TEGEL, H., MULDER, J., ROCKBERG, J., NILSSON, P., SCHWENK, J. M., HAMSTEN, M., VON FEILITZEN, K., FORSBERG, M., PERSSON, L., JOHANSSON, F., ZWAHLEN, M., VON HEIJNE, G., NIELSEN, J. & PONTEN, F. 2015. Proteomics. Tissue-based map of the human proteome. *Science,* 347**,** 1260419.

UK COLORECTAL CANCER SCREENING PILOT GROUP 2004. Results of the first round of a demonstration pilot of screening for colorectal cancer in the United Kingdom. *Bmj,* 329**,** 133.

UMMANNI, R., JUNKER, H., ZIMMERMANN, U., VENZ, S., TELLER, S., GIEBEL, J., SCHARF, C., WOENCKHAUS, C., DOMBROWSKI, F. & WALTHER, R. 2008. Prohibitin identified by proteomic analysis of prostate biopsies distinguishes hyperplasia and cancer. *Cancer Lett,* 266**,** 171-85.

VALADÃO, M., LEAL, R. A., BARBOSA, L. C., CARNEIRO, M. & MUHARRE, R. J. 2010. Perfil dos pacientes portadores de câncer colorretal operados em um hospital geral: necessitamos de um programa de rastreamento acessível e efetivo. *Rev Brás Coloproct,* 30**,** 160-6.

VAN AMERONGEN, R. 2012. Alternative Wnt pathways and receptors. *Cold Spring Harb Perspect Biol,* 4.

VAN AMERONGEN, R., MIKELS, A. & NUSSE, R. 2008. Alternative wnt signaling is initiated by distinct receptors. *Sci Signal,* 1**,** re9.

VAN NGUYEN, S., SKARSTEDT, M., LÖFGREN, S., ZAR, N., ANDERSSON, R. E., LINDH, M., MATUSSEK, A. & DIMBERG, J. 2013. Gene polymorphism of matrix metalloproteinase-12 and -13 and association with colorectal cancer in Swedish patients. *Anticancer Res,* 33**,** 3247-50.

VAN ROSSUM, L. G., VAN RIJN, A. F., LAHEIJ, R. J., VAN OIJEN, M. G., FOCKENS, P., VAN KRIEKEN, H. H., VERBEEK, A. L., JANSEN, J. B. & DEKKER, E. 2008. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. *Gastroenterology,* 135**,** 82-90.

VAN WEEMEN, B. K. & SCHUURS, A. H. 1971. Immunoassay using antigen-enzyme conjugates. *FEBS Lett,* 15**,** 232-236.

VARGHESE, F., BUKHARI, A. B., MALHOTRA, R. & DE, A. 2014. IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. *PLoS One,* 9**,** e96801.

VARMA, M. & JASANI, B. 2005. Diagnostic utility of immunohistochemistry in morphologically difficult prostate cancer: review of current literature. *Histopathology,* 47**,** 1-16.

VICHAI, V. & KIRTIKARA, K. 2006. Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nat Protoc,* 1**,** 1112-6.

VOGEL, C. & MARCOTTE, E. M. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet,* 13**,** 227-32.

VOGELSTEIN, B., FEARON, E. R., HAMILTON, S. R., KERN, S. E., PREISINGER, A. C., LEPPERT, M., NAKAMURA, Y., WHITE, R., SMITS, A. M. & BOS, J. L. 1988. Genetic alterations during colorectal-tumor development. *N Engl J Med,* 319**,** 525-32.

WANG, E. T., SANDBERG, R., LUO, S., KHREBTUKOVA, I., ZHANG, L., MAYR, C., KINGSMORE, S. F., SCHROTH, G. P. & BURGE, C. B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature,* 456**,** 470-6.

WANG, X., CUNNINGHAM, M., ZHANG, X., TOKARZ, S., LARAWAY, B., TROXELL, M. & SEARS, R. C. 2011. Phosphorylation regulates c-Myc's oncogenic activity in the mammary gland. *Cancer Res,* 71**,** 925-36.

WANG, Y., BLANDINO, G., OREN, M. & GIVOL, D. 1998. Induced p53 expression in lung cancer cell line promotes cell senescence and differentially modifies the cytotoxicity of anti-cancer drugs. *Oncogene,* 17**,** 1923-30.

WARBURG, O., WIND, F. & NEGELEIN, E. 1927. THE METABOLISM OF TUMORS IN THE BODY. *J Gen Physiol,* 8**,** 519-30.

WARTHIN, A. S. 1913. Heredity with reference to carcinoma: As shown by the study of the cases examined in the pathological laboratory of the University of Michigan, 1895-1913. *Archives of Internal Medicine,* 12**,** 546.

WATSON, M. M. & SØREIDE, K. 2016. The prognostic yield of biomarkers harvested in chemotherapy-naive stage II colon cancer: can we separate the wheat from the chaff? *Mol Med,* 22. 271-273

WEBBER, E. M., LIN, J. S. & EVELYN, P. W. 2010. Oncotype DX tumor gene expression profiling in stage II colon cancer. Application: prognostic, risk prediction. *PLoS Curr,* 2**,** RRN1177.

WEI, H. & WANG, E. 2011. Electrochemiluminescence of tris(2,2'-bipyridyl)ruthenium and its applications in bioanalysis: a review. *Luminescence,* 26**,** 77-85.

WEINBERG, R. A. 2007. *The biology of cancer,* New York, Garland Science.

WINAWER, S. J., ZAUBER, A. G., HO, M. N., O'BRIEN, M. J., GOTTLIEB, L. S., STERNBERG, S. S., WAYE, J. D., SCHAPIRO, M., BOND, J. H. & PANISH, J. F. 1993. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med,* 329**,** 1977-81.

WITSCH, E., SELA, M. & YARDEN, Y. 2010. Roles for growth factors in cancer progression. *Physiology (Bethesda),* 25**,** 85-101.

WONG, S. C., LO, E. S., CHAN, A. K., LEE, K. C. & HSIAO, W. L. 2003. Nuclear beta catenin as a potential prognostic and diagnostic marker in patients with colorectal cancer from Hong Kong. *Mol Pathol,* 56**,** 347-52.

WONG, S. C., LO, E. S., LEE, K. C., CHAN, J. K. & HSIAO, W. L. 2004. Prognostic and diagnostic significance of beta-catenin nuclear immunostaining in colorectal cancer. *Clin Cancer Res,* 10**,** 1401-8.

WOOD, L. D., PARSONS, D. W., JONES, S., LIN, J., SJÖBLOM, T., LEARY, R. J., SHEN, D., BOCA, S. M., BARBER, T., PTAK, J., SILLIMAN, N., SZABO, S., DEZSO, Z., USTYANKSKY, V., NIKOLSKAYA, T., NIKOLSKY, Y., KARCHIN, R., WILSON, P. A., KAMINKER, J. S., ZHANG, Z., CROSHAW, R., WILLIS, J., DAWSON, D., SHIPITSIN, M., WILLSON, J. K., SUKUMAR, S., POLYAK, K., PARK, B. H., PETHIYAGODA, C. L., PANT, P. V., BALLINGER, D. G., SPARKS, A. B., HARTIGAN, J., SMITH, D. R., SUH, E., PAPADOPOULOS, N., BUCKHAULTS, P., MARKOWITZ, S. D., PARMIGIANI, G., KINZLER, K. W., VELCULESCU, V. E. & VOGELSTEIN, B. 2007. The genomic landscapes of human breast and colorectal cancers. *Science,* 318**,** 1108-13.

WORLD HEALTH ORGANIZATION. 2008. *The global burden of disease: 2004 update.* [Online]. Geneva: World Heath Organization. Available: http://www.who.int/topics/global_burden_of_disease/en/ [Accessed 01 September 2014].

WU, C. C., CHEN, H. C., CHEN, S. J., LIU, H. P., HSIEH, Y. Y., YU, C. J., TANG, R., HSIEH, L. L., YU, J. S. & CHANG, Y. S. 2008. Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. *Proteomics,* 8**,** 316-32.

WU, H., DING, Z., HU, D., SUN, F., DAI, C., XIE, J. & HU, X. 2012a. Central role of lactic acidosis in cancer cell resistance to glucose deprivation-induced cell death. *J Pathol,* 227**,** 189-99.

WU, Q., GOU, Y., WANG, Q., JIN, H., CUI, L., ZHANG, Y., HE, L., WANG, J., NIE, Y., SHI, Y. & FAN, D. 2011. Downregulation of RPL6 by siRNA inhibits proliferation and cell cycle progression of human gastric cancer cell lines. *PLoS One,* 6**,** e26401.

WU, T. F., WU, H., WANG, Y. W., CHANG, T. Y., CHAN, S. H., LIN, Y. P., LIU, H. S. & CHOW, N. H. 2007. Prohibitin in the pathogenesis of transitional cell bladder cancer. *Anticancer Res,* 27**,** 895-900.

WU, X., ZHANG, J., HE, X., WANG, C., LIAN, L., LIU, H., WANG, J. & LAN, P. 2012b. Postoperative adjuvant chemotherapy for stage II colorectal cancer: a systematic review of 12 randomized controlled trials. *J Gastrointest Surg,* 16**,** 646-55.

XIAO, R., SUN, Y., DING, J. H., LIN, S., ROSE, D. W., ROSENFELD, M. G., FU, X. D. & LI, X. 2007. Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol Cell Biol,* 27**,** 5393-402.

XIE, R., CHUNG, J. Y., YLAYA, K., WILLIAMS, R. L., GUERRERO, N., NAKATSUKA, N., BADIE, C. & HEWITT, S. M. 2011. Factors influencing the degradation of archival formalin-fixed paraffin-embedded tissue sections. *J Histochem Cytochem,* 59**,** 356-65.

XIE, W. C., CHAN, M. H., MAK, K. C., CHAN, W. T. & HE, M. 2012. Trends in the incidence of 15 common cancers in Hong Kong, 1983-2008. *Asian Pac J Cancer Prev,* 13**,** 3911-6.

XU, X., KIM, J. E., SUN, P. L., YOO, S. B., KIM, H., JIN, Y. & CHUNG, J. H. 2015. Immunohistochemical demonstration of alteration of beta-catenin during tumor metastasis by different mechanisms according to histology in lung cancer. *Exp Ther Med,* 9**,** 311-318.

XUE, W., ZENDER, L., MIETHING, C., DICKINS, R. A., HERNANDO, E., KRIZHANOVSKY, V., CORDON-CARDO, C. & LOWE, S. W. 2007. Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature,* 445**,** 656-60.

YALOW, R. S. & BERSON, S. A. 1960. Immunoassay of endogenous plasma insulin in man. *J Clin Invest,* 39**,** 1157-75.

YAMADA, T., OSHIMA, T., YOSHIHARA, K., TAMURA, S., KANAZAWA, A., INAGAKI, D., YAMAMOTO, N., SATO, T., FUJII, S., NUMATA, K., KUNISAKI, C., SHIOZAWA, M., MORINAGA, S., AKAIKE, M., RINO, Y., TANAKA, K., MASUDA, M. & IMADA, T. 2010. Overexpression of MMP-13 gene in colorectal cancer with liver metastasis. *Anticancer Res,* 30**,** 2693-9.

YAMADA, Y. & MORI, H. 2007. Multistep carcinogenesis of the colon in Apc(Min/+) mouse. *Cancer Sci,* 98**,** 6-10.

YAMANE, L., SCAPULATEMPO-NETO, C., REIS, R. M. & GUIMARAES, D. P. 2014. Serrated pathway in colorectal carcinogenesis. *World J Gastroenterol,* 20**,** 2634-40.

YAN, C., MENG, X. U. N., LAN, L. I. & YAN-E, G. A. O. 2013. Expression and significance of prohibitin in cervical squamous cell carcinoma. *西安交通大学学报 (医学版) = Journal of Xi'an Jiaotong University (Medical Sciences),* 34**,** 586-588.

YAN, Y., YIN, P., GONG, H., XUE, Y., ZHANG, G., FANG, B., CHEN, Z., LI, Y., YANG, C., HUANG, Z., YANG, X., GE, J. & ZOU, Y. 2016. Nucleosome Assembly Protein 1-Like 1 (Nap1l1) Regulates the Proliferation of Murine Induced Pluripotent Stem Cells. *Cell Physiol Biochem,* 38**,** 340-50.

YANG, B., GAO, J., RAO, Z. & SHEN, Q. 2012. Clinicopathological significance and prognostic value of MMP-13 expression in colorectal cancer. *Scand J Clin Lab Invest,* 72**,** 501-5.

YARDEN, Y. & SLIWKOWSKI, M. X. 2001. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol,* 2**,** 127-37.

YE, J., WU, D., WU, P., CHEN, Z. & HUANG, J. 2014. The cancer stem cell niche: cross talk between cancer stem cells and their microenvironment. *Tumour Biol,* 35**,** 3945-51.

YI, R., LI, Y., WANG, F. L., MIAO, G., QI, R. M. & ZHAO, Y. Y. 2016. MicroRNAs as diagnostic and prognostic biomarkers in colorectal cancer. *World J Gastrointest Oncol,* 8**,** 330-40.

YIU, A. J. & YIU, C. Y. 2016. Biomarkers in Colorectal Cancer. *Anticancer Res,* 36**,** 1093-102.

YONISH-ROUACH, E., RESNITZKY, D., LOTEM, J., SACHS, L., KIMCHI, A. & OREN, M. 1991. Wild-type p53 induces apoptosis of myeloid leukaemic cells that is inhibited by interleukin-6. *Nature,* 352**,** 345-7.

YOON, Y. S., YU, C. S., KIM, T. W., KIM, J. H., JANG, S. J., CHO, D. H., ROH, S. A. & KIM, J. C. 2011. Mismatch repair status in sporadic colorectal cancer: immunohistochemistry and microsatellite instability analyses. *J Gastroenterol Hepatol,* 26**,** 1733-9.

YOSHIDA, N., KINUGASA, T., OHSHIMA, K., YUGE, K., OHCHI, T., FUJINO, S., SHIRAIWA, S., KATAGIRI, M. & AKAGI, Y. 2015. Analysis of Wnt and beta-catenin Expression in Advanced Colorectal Cancer. *Anticancer Res,* 35**,** 4403-10.

ZAHA, D. C. 2014. Significance of immunohistochemistry in breast cancer. *World J Clin Oncol,* 5**,** 382-92.

ZAUBER, A. G., WINAWER, S. J., O'BRIEN, M. J., LANSDORP-VOGELAAR, I., VAN BALLEGOOIJEN, M., HANKEY, B. F., SHI, W., BOND, J. H., SCHAPIRO, M., PANISH, J. F., STEWART, E. T. & WAYE, J. D. 2012. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med,* 366**,** 687-96.

ZELLER, K. I., ZHAO, X., LEE, C. W., CHIU, K. P., YAO, F., YUSTEIN, J. T., OOI, H. S., ORLOV, Y. L., SHAHAB, A., YONG, H. C., FU, Y., WENG, Z., KUZNETSOV, V. A., SUNG, W. K., RUAN, Y., DANG, C. V. & WEI, C. L. 2006. Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proc Natl Acad Sci U S A,* 103**,** 17834-9.

ZHANG, B., CAO, X., LIU, Y., CAO, W., ZHANG, F., ZHANG, S., LI, H., NING, L., FU, L., NIU, Y., NIU, R., SUN, B. & HAO, X. 2008. Tumor-derived matrix metalloproteinase-13 (MMP-13) correlates with poor prognoses of invasive breast cancer. *BMC Cancer,* 8**,** 83.

ZHANG, N., KAUR, R., AKHTER, S. & LEGERSKI, R. J. 2009. Cdc5L interacts with ATR and is required for the S-phase cell-cycle checkpoint. *EMBO Rep,* 10**,** 1029-35.

ZHANG, Z., WANG, M., ZHOU, L., FENG, X., CHENG, J., YU, Y., GONG, Y., ZHU, Y., LI, C., TIAN, L. & HUANG, Q. 2015a. Increased HMGB1 and cleaved caspase-3 stimulate the proliferation of tumor cells and are correlated with the poor prognosis in colorectal cancer. *J Exp Clin Cancer Res,* 34**,** 51.

ZHANG, Z., ZHU, S., YANG, Y., MA, X. & GUO, S. 2015b. Matrix metalloproteinase-12 expression is increased in cutaneous melanoma and associated with tumor aggressiveness. *Tumour Biol,* 36**,** 8593-600.

ZHENG, J., CHU, D., WANG, D., ZHU, Y., ZHANG, X., JI, G., ZHAO, H., WU, G., DU, J. & ZHAO, Q. 2013. Matrix metalloproteinase-12 is associated with overall survival in Chinese patients with gastric cancer. *J Surg Oncol,* 107**,** 746-51.

ZHOU, T. B. & QIN, Y. H. 2013. Signaling pathways of prohibitin and its role in diseases. *J Recept Signal Transduct Res,* 33**,** 28-36.

ZUCKER, S. & VACIRCA, J. 2004. Role of matrix metalloproteinases (MMPs) in colorectal cancer. *Cancer Metastasis Rev,* 23**,** 101-17.

ZYADA, M. M. & SHAMAA, A. A. 2008. Is collagenase-3 (MMP-13) expression in chondrosarcoma of the jaws a true marker for tumor aggressiveness? *Diagn Pathol,* 3**,** 26.

9. APPENDICES

9.1. Abstract published in scientific event annals

Queiroz, CJS; Ydy, LRA; Miyajima, F; Jenkins, JR. Immunohistochemical validation of novel candidate biomarkers for colorectal cancer. J Clin Oncol 33, 2015 (suppl; abstr e22251)

Event: 2015 American Society of Clinical Oncology (ASCO) Annual Meeting, Chicago, USA, May 2015

Source: http://meetinglibrary.asco.org/content/145151-156